Spring 5-15-2016

No-slip Billiards

Christopher Lee Cox
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Mathematics

Dissertation Examination Committee:
Renato Feres, Chair
Tim Chumley
John McCarthy
Xiang Tang
Mladen Victor Wickerhauser
Hongkun Zhang

No-slip Billiards

by

Christopher Lee Cox

A dissertation presented to the
Graduate School of Arts & Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2016
St. Louis, Missouri

Table of Contents

List of Figures **iii**

Acknowledgments **v**

Abstract **ix**

1 Introduction **1**

2 The No-slip Collision Model **14**
2.1 Collisions conserving energy and momentum 14
2.2 Examples 17

3 A differential geometric framework for rigid body collisions in $\mathbb{R}^n$ **23**
3.1 Statements of the main results 25
3.2 The Euclidean group and its Lie algebra 40
3.3 Newtonian mechanics of rigid bodies 45
3.4 Kinematics of two rigid bodies 55
3.5 Collision maps 65

4 The Dynamics of No-slip Billiards **69**
4.1 Parallel Boundaries 70
4.2 Open wedges 75
4.3 Circles 82

5 Ergodic No-Slip Billiards **85**
5.1 The invariance of the no-slip map 85
5.2 Ergodic considerations 90

Bibliography **96**

A General Billiards (1.03) **98**

List of Figures

1.1 A no-slip collision 1
1.2 No-slip strip 2
1.3 Equilateral triangle 3
1.4 Geometric description of the collision map for no-slip billiards in dimension 2. 6
1.5 Four no-slip wedges 7
1.6 Specular and no-slip circles 8
1.7 No-slip stadium and Sinai billiards 9
1.8 Reduced phase space for no-slip billiards in two dimensions 10
1.9 Hexagonal no-slip velocity phase space 11
1.10 KAM islands 12

2.1 Combinatorial possibilities for trajectories of the equilateral triangle 19
2.2 Random rough rank 20
2.3 Half smooth, half no-slip disk 21
2.4 3D parallel plates 21
2.5 3D plates varying rough rank 22

3.1 Rigid body configuration 26
3.2 A billiard system 37
3.3 The Ψ map 56
3.4 Avoiding pathological configurations 57
3.5 Situations for which Proposition 19 applies. 60

4.1 2D No-slip strip view in three dimensions 70
4.2 Astroid of contact for the no-slip strip 72
4.3 Axis of periodicity 75
4.4 Axis of periodicity for $\frac{\pi}{2}$ wedge 76
4.5 Wedge angle θ and velocity rotation angle α 78
4.6 Period eight wedge 78
4.7 Escape and non-escape sphere for the no-slip wedge 80
4.8 Graph of product form iterations for wedge 81
4.9 No-slip circle caustics 82
4.10 No-slip circle dynamics 83

5.1 The canonical measure 86

5.2 Experiment to illustrate invariance of the billiard measure for a rectangular table. . 88
5.3 Isosceles phase space 90
5.4 Pentagon perturbations 91
5.5 Focusing and defocusing 92
5.6 No-slip billiard phase portraits for standard ergodic tables 93
5.7 No-slip sinai at the elliptic threshold 94
5.8 No-slip sinai at the elliptic threshold wedge 95

Acknowledgments

If I were to properly thank all the people who have helped me over the just last five years, the acknowledgments section would swell from its current eye-raising three pages to an outrageously inappropriate length. Nonetheless, I think it is worth the effort to thank many people individually. As indebted as I am, this will necessarily result in the omission of a few worthy names–to anyone in that category reading this, I beg your forgiveness. Please consider that while my gratitude necessarily exceeds that of a typical doctoral student, the frequency of my memory lapses does as well.

First and foremost, I would like to thank Ben Passer for creating the template for this dissertation, one of the many ways my fellow graduate students have helped me along this journey. When I returned to the world of higher math in the form of qualifying courses, working through homework exercises with Dave Meyer, Ryan Keast, and Joey Palmer helped me survive the transition. In the subsequent years, while at the office I enjoyed many profitable and enjoyable conversations with, just to name a few, Brady Rocks, Wei Deng, my second favorite Romanian Alex Papiu, Meredith Sargent, Marie Jose Saad, Luis Garcia, and my second favorite Brazilian Genival Fernandes da Silva, Jr.

Too many years ago to count, my lifelong love of math was born under the guidance and tutelage of Mr. Gary Millen, who introduced me to soap bubbles as math.[1] Not quite so many years ago, I had the privilege of being an undergraduate math major at Williams College, and while the list of great teachers I encountered there could fill a second acknowledgements section, I would be remiss if I did not thank my lifelong mentor Frank Morgan and friend Tom Garrity, neither of whom tried very hard at all to talk me out of going back to graduate school when I visited in 2010.

[1] In my earliest extant mathematical paper, I use an applied approach to demonstrate that planar soap bubbles meet an an angle of 119.5 degrees.

My colleagues at the Illinois Central College Math Department, including Deanna Welsch, Fereja Tahir, Yolanda Rush, Pat Ward, and Elizabeth "Liz" Carrico, have both my gratitude and continued respect for their tireless work on the front lines of math education.

Having entered the graduate program with extensive teaching experience, I was doubtful that attending a teaching seminar was a good use of my time. In retrospect my time could not have been better spent, and I am grateful to the insights and encouragement I received during that three semester experience from Danny Zadkovic, Cheng Chu, Joey, and Blake Thornton. Throughout my time as a teaching assistant, Blake continued to be a mentor, colleague, and friend; few people have taught me more about teaching or won longer SUP races.

I am especially grateful to Tim Chumley, first for his encouragement during my early days as a returning graduate student, and more recently for inviting me to be a part of the American Institute of Mathematics Stochastic Thermodynamics and Random Billiards SQuaRE. I look forward to continuing to work with Tim and the other members of the group, Scott Cook, Matt Wallace, and Hongkun Zhang, later this year and in 2017.

The Washington University Math Department is full of people deeply committed to student success. I am especially indebted to those who helped me through my qualifying courses, Matt Kerr, Richard Rochberg, Rachel Roberts, whose detailed lectures were an essential part of my early success, and John McCarthy, whose deep commitment to understanding math is matched by his commitment to helping others understand. To Victor Wickerhauser I am grateful for the canoe lessons, and in particular for his insistence that there was no need to launch an investigation about who it was who paddled on the wrong side going through the rapids and tipped the canoe, drenching everyone.

Besides my advisor, no one else was more instrumental in my success at Washington University than Xiang Tang. He was my assigned advisor when I arrived and helped chart my course the first two years. He supported me over two summers and through three semesters

of reading in Symplectic Geometry. Taking a course from Xiang is not easy, and taking a reading course will expose a student to intense questioning at times. But, as one of my fellow students observed, "after a while you realize he does it because he really cares about you." No doubt that is also the reason he signed me up to speak every semester in the Topology and Geometry Seminar, advancing my progress far more rapidly then would have been the case were I required to volunteer to speak.

In November 2010 I had left a rewarding though demanding career teaching at a community college and was in the process of finalizing a list of doctoral programs in mathematics. That list did not include Washington University, but when in the process of due diligence I came across the Math Department homepage and read David Wright's welcome, it seemed worthy of consideration. By the time I arrived for a visit the following April, I had decided to accept the admission offer. Although I needed no further convincing, the enthusiasm Mary Ann Stenner showed in introducing me to the department infused me with energy and convinced me that I had made the right decision. Having arrived at the other side, I am more convinced than ever. The support offered from the Washington University Math Department makes it an excellent place to earn a math Ph.d.

If I were charged with constructing an ideal math professor, I would insist on three qualities: a driving curiosity, a deep humility, and a relentless work ethic. I have been lucky enough to work with someone amply blessed with all three. Many years ago, when I was an undergraduate, I was warned that success or failure in graduate school would hinge more than anything else on finding a good advisor. I could not have hoped to find a better advisor than Renato Feres.

Christopher Lee Cox

Washington University in St. Louis

May 2016

Dedicated to Lyta.

ABSTRACT OF THE DISSERTATION

No-slip Billiards

by

Christopher Lee Cox

Doctor of Philosophy in Mathematics

Washington University in St. Louis, 2016

Professor Renato Feres, Chair

We investigate the dynamics of no-slip billiards, a model in which small rotating disks may exchange linear and angular momentum at collisions with the boundary. A general theory of rigid body collisions in $\mathbb{R}^n$ is developed, which returns the known dimension two model as a special case but generalizes to higher dimensions. We give new results on periodicity and boundedness of orbits which suggest that a class of billiards (including all polygons) is not ergodic. Computer generated phase portraits demonstrate non-ergodic features, suggesting chaotic no-slip billiards cannot easily be constructed using the common techniques for generating chaos in standard billiards. However, Sinai type dispersing billiards, which are always ergodic in the case of standard billiards, appear to be ergodic above a certain curvature threshold.

Chapter 1

Introduction

No-slip billiards are a type of billiard dynamical system based upon a model in which linear and angular momentum of a hard spherical particle, moving in an n-dimensional Euclidean domain, may be exchanged on collisions at the boundary with total energy conserved, as indicated in the diagram of Figure 1.1. The standard billiard model is based on a specular collision in which the angle of reflection equals the angle of incidence from the boundary. The origins of billiard dynamics can be traced back to Jacques Hadamard [17] at the end of the nineteenth century, and in the half century since Yakov Sinai's [23, 24] seminal insights and the subsequent work of his students Bunimovich [8] and Chernov[9], standard billiards have been and continue to be extensively studied. In contrast, though the no-slip model may be seen to arise as a second ideal collision on equal footing with the specular collision model, research of no-slip billiards has been limited. Our purpose is to begin a systematic study of no-slip billiards, looking at the geometric foundation, the dynamics, and the ergodic properties.

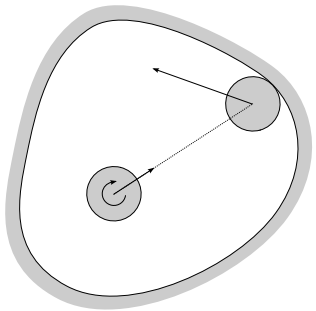

Figure 1.1: In a no-slip billiard system in dimension 2 a form of non-dissipative friction at collisions causes linear and rotational velocities to be partially exchanged.

Chapter 2 briefly introduces the no-slip model and then demonstrates its application with several examples. While elements of the no-slip model appear in the physics literature in the context of "rough" conservative collisions [16][28], the first systematic development was by Broomhead and Gutkin [5]. They introduced no-slip collisions in two dimensions as a new model of a gas, in which "the spheres interact with each other and with the container walls" without slipping, in such a way that the impact "conserves the total energy of the system, but mixes the tangential velocity components with the angular velocities of colliding spheres." They demonstrate that in two dimensions energy may be conserved only by the specular collisions of the traditional gas model, well-known from standard billiards, or the unique conservative alternative of no-slip collisions, and as an application they describe the following fundamental example.

Example 1. *(The no-slip infinite strip)* In contrast to the specular analogue, the dimension two, no-slip billiard between two (infinite) parallel boundaries will have bounded orbits for any initial rotational and linear velocities, unless the velocity component perpendicular to the boundary is zero. Gutkin and Broomhead showed the boundedness using complex coordinates for the phase space and showing that the series giving the horizontal displacement after n collisions is bounded.

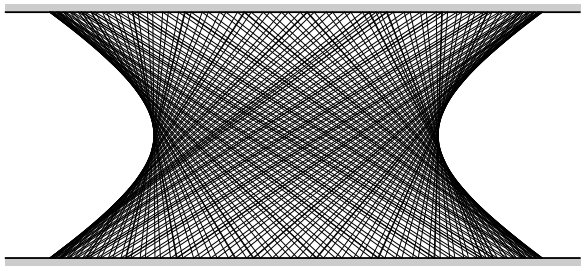

Figure 1.2: No-slip collisions between planar parallel boundaries have maximal displacement $\sqrt{\frac{3}{2}\left(\frac{1}{(\dot{x}_2)^2}-1\right)}$, corresponding to the length of the base of the region in the above figure containing the trajectory. Here $\dot{x}_2$ is the (constant) vertical velocity and the particle is a disk of uniform mass.

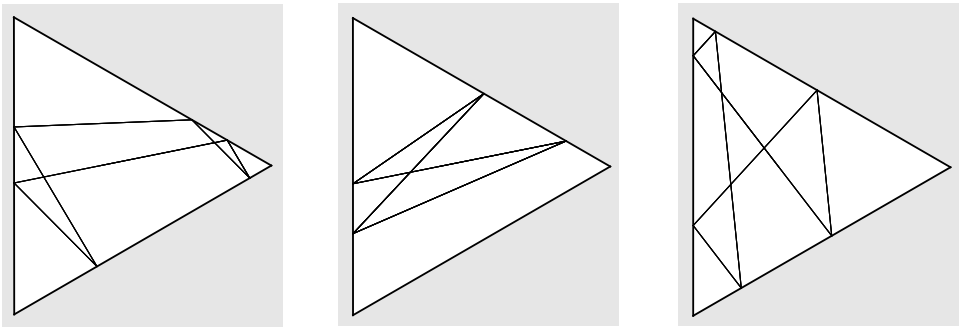

Figure 1.3: The no-slip equilateral triangle is periodic for all initial conditions. Orbits may be of the three types above or their degeneracies.

The broader implications of the boundedness of this example were not pursued in [5], but analytical and numerical observations suggest this property holds in much greater generality and is at the heart of a very general non-ergodicity result for this class of billiard systems.

The last section of Chapter 2 presents several examples which show interesting features of the no-slip model but which are not pursued in detail at this time. For example, from the viewpoint of allowing each boundary point to be either "rough" or "smooth", we revisit the no-slip strip of Example 1 and consider a hybrid model where both types of collisions are possible. If the collision is randomly assigned, the system is merely a one dimensional random walk. More interestingly, if the disk is given rough and smooth sides and the rotational position is tracked so that the collision is assigned by the disk's position, the system will exhibit a type of pseudoperiodic behavior which changes according to scale. Another interesting example is the no-slip equilateral triangle billiard, as in Figure 1.3. Regardless of what initial parameters are chosen, the system will be periodic, with period either four, six, or degenerate versions of these with period two or three.[1]

Chapter 3 fills in the background found in joint work with Feres and Ward [13]. We

[1]Note that the rotation adds a third dimension to the two spacial dimensions. Often, if there is no loss of information, the given figures will be the planar projection, as in Figure 1.3. Also, without any essential alteration of the theory we may use the orbit of the center of mass, with boundaries accordingly adjusted by distance R.

introduce a broad framework for collisions of rigid bodies in $\mathbb{R}^n$ satisfying natural physical requirements which yields as a special case the no-slip model in $\mathbb{R}^2$, but which easily extends to higher dimensions and could be used to consider broader classes of collision models. The starting point is considering collisions as maps on the tangent space of the configuration manifold at boundary points.

Definition 1 (Strict collision maps)**.** Let M denote the configuration manifold of two rigid bodies in $\mathbb{R}^n$ having smooth boundaries. We endow M with the Riemannian metric whose quadratic form gives the system's kinetic energy function and assume that points $q \in \partial M$ represent configurations in which the bodies have a single contact point. At a boundary point q of M where ∂M is differentiable, we define a *collision map* as a linear map $\mathcal{C} : T_qM \to T_qM$ that sends vectors pointing out of M into vectors pointing inward. We say that $\mathcal{C}$ is a *strict collision map* at q if

1. Energy is conserved. That is, $\mathcal{C}$ is an orthogonal linear map;

2. Linear and angular momentum are conserved in the unconstrained motion. (This property, expressed in terms of invariance of a momentum map, may amount to no restriction at all when one of the bodies, representing the billiard table, is assumed fixed in place and the subgroup of Euclidean symmetries of the whole system is trivial);

3. Time reversibility. This amounts to $\mathcal{C}$ being a linear involution;

4. Impulse forces at collision are applied only at the single point of contact. (See [13] for an elaboration and geometric interpretation of this property, which may be regarded as a generalized momentum conservation law that is generally non-trivial and highly restrictive.)

We give a classification of collisions maps, including the following theorem proved in Section 3.1.

Theorem 2. At each boundary point of the configuration manifold of the system of two rigid bodies, assuming the boundary is differentiable at that point, the set of all strict collision maps can be expressed as the disjoint union of orthogonal Grassmannian manifolds $Gr(k, n-1)$, $k = 0, \ldots, n-1$, of all k dimensional planes in $\mathbb{R}^{n-1}$. In particular, when $n = 2$, the set of strict collision maps is a two-point set consisting of the specular reflection and the no-slip collision.

This theorem implies that for dimension $n \geq 3$, there exists a spectrum of possible choices in contrast to the two choices in dimension 2. One might require specular collisions for all directions at all boundary points, in which case the model is simply that which is used for higher dimensional standard billiards, or one might assign no-slip collisions everywhere. Alternatively, though, at each boundary point the requirements of a strict collision will also be satisfied by hybrid models assigning some smooth and some rough directions. To describe these cases we will introduce the notion of "rough rank", discussed informally in Section 2.2 and formalized in Section 3.1.4.

We are most interested in applying this general theory of rigid body collisions to billiard systems, generally understood to consist of two rigid bodies, one of which is fixed in place and called the *billiard table*. The focus is then on the motion of the second, referred to as the *billiard particle*. For the system to be fully specified it is necessary to impose boundary conditions. From our perspective, this amounts to assigning a collision map $\mathcal{C}_q$ to each boundary configuration $q \in \partial M$ from among those in the moduli of collision maps described by the above theorem. The geometric description of a no-slip collision of a billiard particle at a wall is explained in Figure 1.4. In addition to the standard reflection it involves a rotation by the special angle β, which is the same for all table shapes under the assumption that the particle is a disk of uniform density.

In Chapter 4 we look more closely at the dynamics of no-slip billiards, primarily focusing two spacial dimensions. We begin by returning to Example 1, giving an alternate proof of boundedness which gives the bound mentioned in Figure 1.2 and which extends naturally to

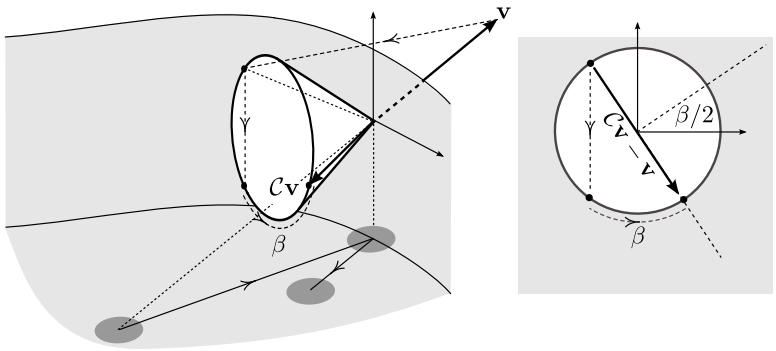

Figure 1.4: Geometric description of the collision map for no-slip billiards in dimension 2. The vertical axis, perpendicular to the billiard table, represents the angle of rotation of the disk (linearly scaled so that total kinetic energy becomes the Euclidean norm). The total (linear and angular) velocity of the billiard particle after collision, $\mathcal{C}\mathbf{v}$, lies in the cone determined by the incoming velocity $\mathbf{v}$ as indicated in the figure. To obtain $\mathcal{C}\mathbf{v}$, first reflect $\mathbf{v}$ back through the vertical axis so as to point into the 3-dimensional configuration space, change the sign of its rotational component, and finally rotate the resulting vector by the angle β such that $\cos\beta = 1/3$ and $\sin\beta = 2\sqrt{2}/3$. This β corresponds to a disk with uniform mass distribution and more generally would depend on the moment of inertia.

a proof that no-slip billiards in dimension three between planes are bounded. We investigate periodicity of no-slip billiards, beginning by showing that the strip will not be periodic except for the trivial period two case involving no rotational or horizontal velocity.

One driving question in the study of standard billiards is whether for a given table the billiard map is *ergodic*, that is, whether every set which is invariant under the billiard flow has measure 0 or 1. While this topic is deferred until the last chapter, it serves as one lens through which we consider the dynamics. For example, a feature of no-slip billiards that may preclude ergodicity is an *axis of periodicity*, a certain type of period two trajectory that will be seen to be ubiquitous. In contrast to the hyperbolic periodic points that may exist in standard ergodic billiards, these axes appear to occur as the center of small invariant regions near elliptic periodic points. These axes occur in a large class of billiards including all polygons. This behavior demonstrates one marked difference between the dynamics of no-slip billiards and those of standard billiards. For example, generic polygons (having angles of irrational multiples of π) for standard billiards are ergodic [18], but no-slip billiards on polygons appear

to be non-ergodic. Fundamental to understanding this behavior is the following example, investigated more fully in Section 4.2.

Example 2. *(The no-slip wedge)* Consider the no-slip billiard table consisting of the unbounded region between two rays meeting at angle θ. An informal numerical survey of wedge systems reveals that for most angles $\theta \in (0, \pi)$ the behavior is similar to the bounded, nonperiodic dynamics of the no-slip strip (Figure 1.5). However, certain isolated θ have periodic orbits which are stable in the sense of persisting when the initial conditions are altered, to any extent not inducing a direct exit from the wedge.

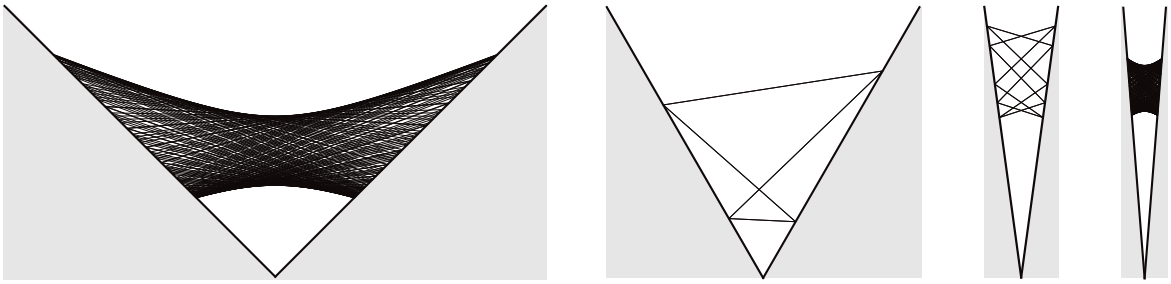

Figure 1.5: For most angles, no-slip open wedges give nonperiodic orbits (far left and far right), but $\theta = \frac{\pi}{3}$ (center left) and $\theta \approx .2709$ (center right) give period four and period ten orbits which persist when initial velocities are changed.

This persistent periodicity, it turns out, is more common than is immediately apparent numerically, and in fact can be precisely related to the wedge angle. Consequences of this fact and another useful result gleaned from the wedge example are given in the following theorem, proved in Section 4.2.

Theorem 3. For a no-slip billiard wedge of angle $\theta \in (0, \pi)$, let x_0 be the rotational axis, x_2 the direction of the wedge bisector, and x_1 the perpendicular spacial direction.

i There exists a periodic axis, a direction in which all trajectories are periodic. Specifically, for velocity $(\dot{x_0}, \dot{x_1}, \dot{x_2})$, the orbit will be periodic whenever

$$\frac{\dot{x_0}}{\dot{x_1}} = -\sqrt{2} \sin \frac{\theta}{2}.$$

ii For any $n \in \mathbb{Z}^+$, wedge angle θ_n can be chosen so that all non-escape velocities yield $2n$-periodic orbits. Furthermore, the set of all such θ_n is dense in $(0, \pi)$.

iii The angle ψ between the velocity and the axis of periodicity is invariant throughout an orbit, remaining unchanged after collisions.

The final section of Chapter 4 looks at no-slip circle billiards. While standard circular billiards are characterized by a circular caustic, the no-slip circle has a caustic consisting of two circular components (Figure 1.6).

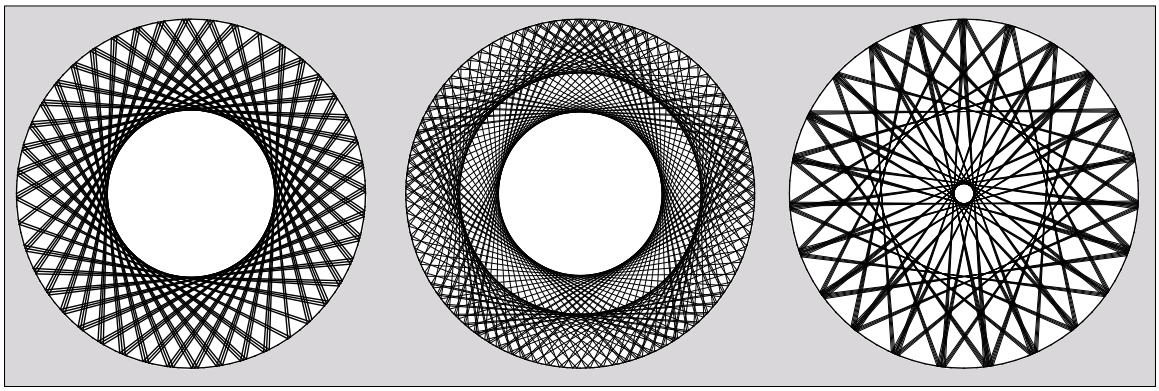

Figure 1.6: The standard circular billiard with circular caustic (left) along with two no-slip circles (middle and right) showing a caustic consisting of two circular components.

In Chapter 5 we look at ergodic theory as it pertains to no-slip billiards. We begin by establishing the invariance of the two dimensional no-slip map, a prerequisite for the remainder of the chapter, and considering the invariance in higher dimensions. A sufficient condition for the invariance of the Liouville measure is that the field of collision maps $q \mapsto \mathcal{C}_q$ be parallel with respect to the Levi-Civita connection on ∂M associated to the kinetic energy Riemannian metric. This condition is satisfied for no-slip billiards in the plane, giving the following theorem.

Theorem 5. The canonical billiard measure on the boundary of the phase space of a planar no-slip billiard is invariant under the billiard map.

In higher dimensions, hybrid billiards (see Example 7) which are not fully rough or fully smooth open the possibility of a non-parallel field of collision maps, in which case the measure might not be invariant.

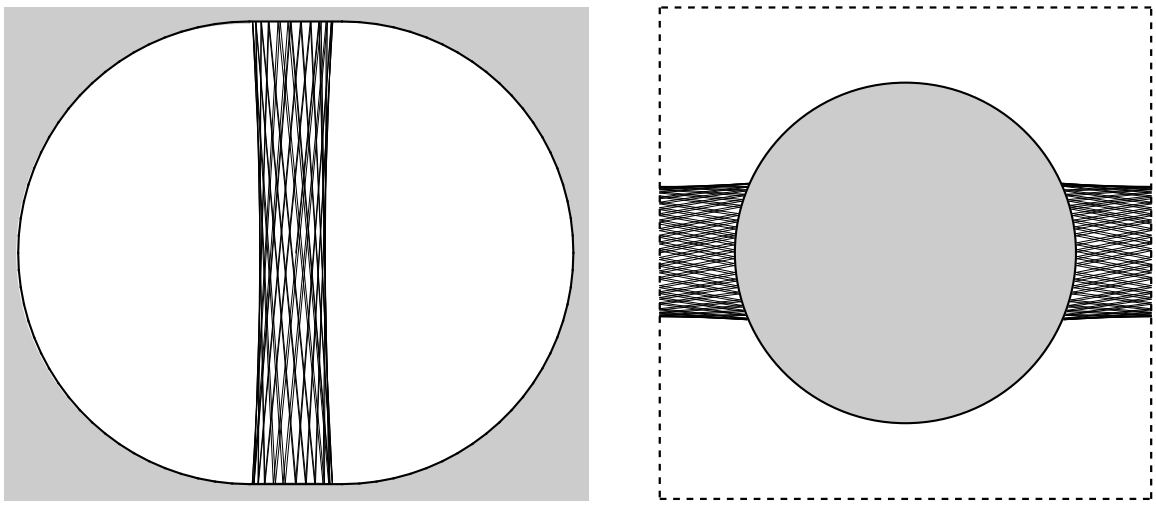

Figure 1.7: The two primary techniques of generating chaotic behavior in standard billiards, defocusing and dispersing, do not immediately translate to no-slip billiards. The Bunimovich stadium (left) is not ergodic, as the flat edges will always produce a positive measure region of phase space with bounded orbits. Similar bounded regions appear in dispersing billiards with concave boundaries for certain parameters, like the Sinai type example of a disk on a torus (right).

Focusing on the dimension 2 case, the remainder of Chapter 5 considers the question of whether ergodic no-slip billiards can be constructed. In standard billiards, two well-known techniques for generating chaotic dynamics are the defocusing and dispersing mechanisms. The stadium of Bunimovich [6], the most well-known example of chaotic focusing, consists of two half circles connected by flat segments. The corresponding no-slip billiard, however, fails to be ergodic, as by [5] any arbitrarily small flat strip will have bounded orbits for all trajectories in a neighborhood of positive measure near the vertical trajectory (Figure 1.7, left). While no such elementary argument can be used to show that the dispersing examples of standard billiards are not ergodic in the no-slip case, Wojtkowski [29] considered periodic orbits with zero tangential velocity and showed that for small curvature stable elliptic points

exist (Figure 1.7, right).

The periodic orbit in the no-slip strip can be viewed as a special case of a periodic axis, and the discussion of the wedge properties along the axis of periodicity in Chapter 4 are supported by the fact that the associated periodic points are elliptic, suggesting the following holds.

Conjecture 1. If a billiard table has an axis of periodicity, with both contact points in flat neighborhoods of the boundary, then the no-slip billiards map is non-ergodic. In particular, for any polygonal table the no-slip map is not ergodic.

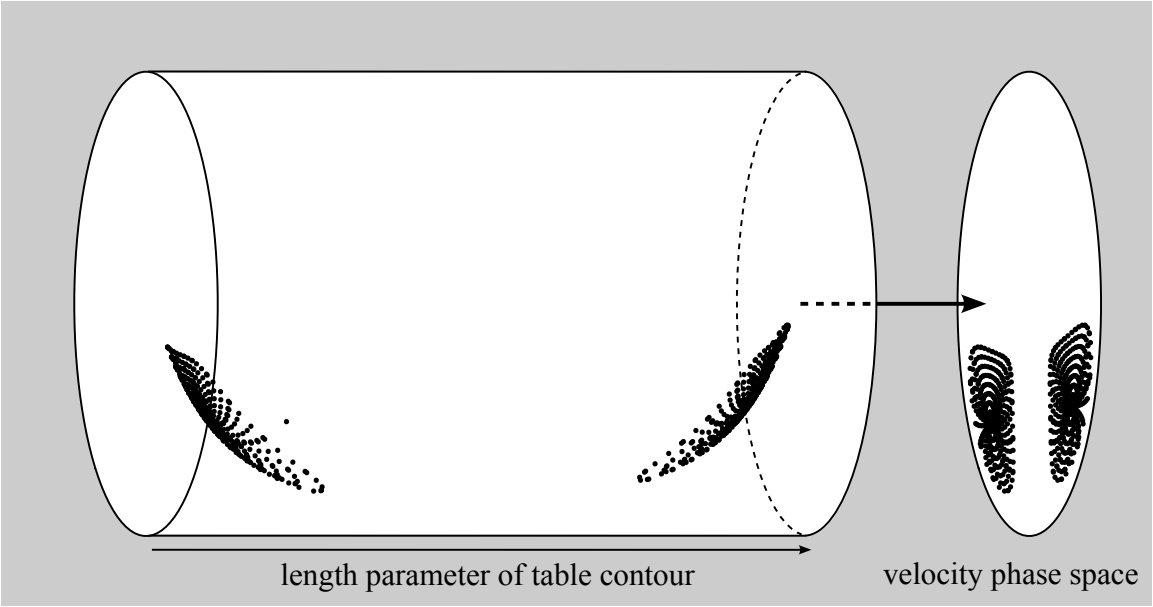

Figure 1.8: Reduced phase space for no-slip billiards in two dimensions is a solid torus, with the horizontal parameter being the position on the boundary. Disk cross sections give the possible velocities at that boundary point. Many of the diagrams shown below represent the projection from the solid torus to a single disk by ignoring the length parameter along the boundary of the table. We call this disk the *velocity phase space*.

In the final section, we consider numerical evidence from phase portraits of no-slip billiards. Note that no-slip billiards in dimension 2 correspond to configuration spaces M of dimension 3, whose points are parametrized by the location of the center of the disk-particle and its

angle of rotation. The boundary of M is then a two-dimensional, piecewise smooth manifold, and the phase space is a 4-dimensional manifold whose points are pairs (q, v) in which $q \in \partial M$ and v may be taken to lie in a hemisphere (whose radius is determined by the conserved kinetic energy) about the normal vector to ∂M at q pointing into M. We identify this hemisphere with the disk of same radius in the tangent space to ∂M at q under the natural orthogonal projection. Thus the phase space of our billiards, at least in the case of bounded (and connected) billiard tables, will typically be homeomorphic to the Cartesian product of a 2-torus and a disk. A simplification results by noting that the angle of rotation (but certainly not the speed of rotation) is typically immaterial. Formally this means that we may for many purposes consider the *reduced phase space*, defined as the quotient of the 4-dimensional phase space by the natural action of the rotation group $SO(2)$. The resulting 3-dimensional solid torus is indicated in the diagram of Figure 1.8. In many cases key features may be gleaned from the projection to the velocity components, which we refer to as the velocity phase portrait.

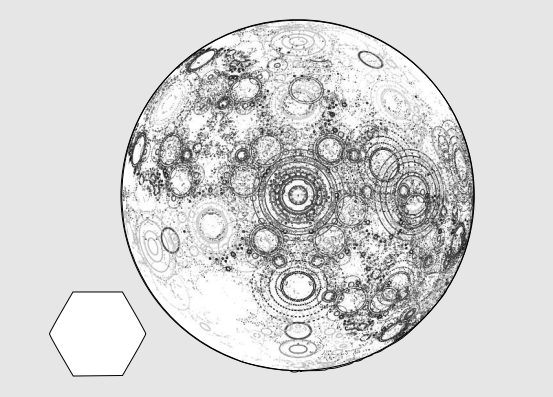

Figure 1.9: The velocity phase portrait is the projection of the 3-dimensional reduced phase portrait as indicated in Figure 1.8. Here orbits of the billiard map for the hexagon no-slip billiard (with table on the left) are shown projected to the velocity phase space.

We conclude with an investigation of phase portraits with an eye towards ergodic candidates. Elliptic periodic points precluding ergodicity will be in evidence as concentric ellipses

in the velocity phase portrait, as in the regular hexagon of Figure 1.9. Indeed, for polygonal tables, phase portraits suggest that there is no positive measure ergodic component. Tables corresponding to known chaotic standard billiards are considered. Many of them appear to display behavior qualitatively similar to invariant KAM islands with surrounding ergodic seas, a much studied and not completely understood phenomenon sometimes found in Hamiltonian dynamical systems indicative of an intermediate system which is neither integrable nor fully chaotic [8]. This behavior has been observed in standard billiards in particular, including including the illustrative mushrooms of Bunimovich [7] and the moon billiards of Correia and Zhang [12]. Figure 1.10 shows the examples of a standard billiard which is a moon variation and a no-slip mushroom.

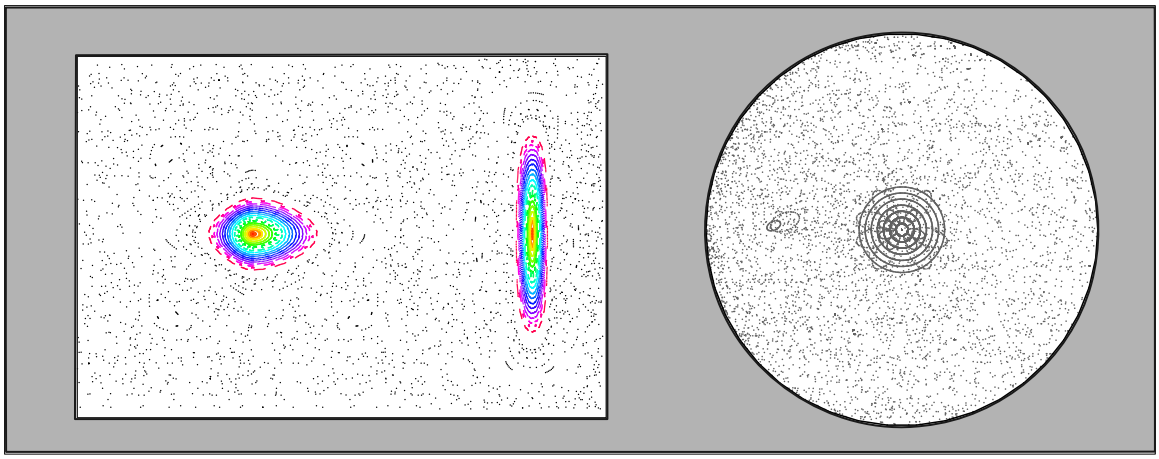

Figure 1.10: The phase portrait of a moon-type standard billiard (left) and the velocity phase portrait of a no-slip mushroom billiard (right). Note the overlap at the center of the mushroom portrait, resulting from projection.

While dispersing no-slip billiards are not generally ergodic, Wojtkowski [29] showed that the elliptic periodic point in evidence in Figure 1.7 above disappears for sufficiently small radii. Simulations confirm this result, showing a rapid dispersal of the structure at the predicted threshold. Further experiments looking at periodic points between non-parallel tangents suggest that the curvature threshold is higher but still exists, and phase portraits

give further evidence that this no-slip billiard on a torus, with a single dispersing disk of small radius, is an excellent candidate for ergodicity. Verifying this analytically would seem to be a high priority for the next step in the study of no-slip ergodic billiards.

Throughout we have relied on simulations to verify and inform the analytical work. Most of these were written using SageMath (System for Algebra and Geometry Experimentation), mathematical software with a python-like syntax. For anyone interested in a more detailed look at this aspect of the work, the code for General Billiards, a broad application for creating both standard and no-slip billiard tables, is given in the appendix.

Chapter 2

The No-slip Collision Model

In section 2.1 we give an elementary derivation of the no-slip collision model, using only the equations for conservation of momentum and energy. A more general derivation giving the broader context will be taken up in Chapter 3. We review connected references in the physical literature, which are few to the best of our knowledge and in some cases simplify the model, but which do provide a glimpse of the connection between the ideal models and actual physical experiments. The last section of this chapter gives several applications of the model.

2.1 Collisions conserving energy and momentum

The no-slip collision model arises naturally as an alternative to the specular model under the physical requirements of conservation of linear and angular momentum and conservation of energy. Broomhead and Gutkin [5] give a detailed development of the two dimensional case in 1993, while a three dimensional physical version of the model are mentioned by Cross ([14], 2002) in relation to tennis balls and Garwin ([16], 1969) in relation to "ultraelastic" *Super-Balls©*.[1]

Consider the motion in two dimensions of a disk of mass m and radius R which bounces between the (horizontal) floor and ceiling while it rotates. The vertical component of the velocity reverses at each collision, while in the tangential direction a friction force F is exerted for a brief interval Δt. Letting $v_{\pm}$ and $\omega_{\pm}$ represent the pre- and post-collision tangential and

[1]Registered by Wham-O Corporation, 835 E. El Monte St., San Gabriel, Calif. 91776

rotational velocities, the conservation of momentum requires

$$m(v_+ - v_-) = F\Delta t \tag{2.1.1}$$

and the conservation of angular momentum

$$\frac{I}{R}(\omega_+ - \omega_-) = -F\Delta t, \tag{2.1.2}$$

where I is the moment of inertia. By conservation of energy,

$$\frac{1}{2}m{v_+}^2 + \frac{1}{2}I{\omega_+}^2 = \frac{1}{2}m{v_-}^2 + \frac{1}{2}I{\omega_-}^2. \tag{2.1.3}$$

Combining the momentum equations while factoring and simplifying the energy equation yields the system

$$\begin{cases} m(v_+ - v_-) = \frac{I}{R}(\omega_- - \omega_+) \\ m(v_+ - v_-)(v_+ + v_-) = -I(\omega_+ - \omega_-)(\omega_+ + \omega_-) \end{cases} \tag{2.1.4}$$

with two solutions. This is the sense in which the no-slip model might be considered an alternative ideal model. The standard specular reflection is described by the solution $v_+ = v_-$, $\omega_+ = \omega_-$, and $F\Delta t = 0$. (Physically, this implies there is no tangential force, suggesting the "smooth" description.) Alternatively, the solution corresponding to the no-slip collision is given by

$$\begin{aligned} v_+ &= \left(1 - \frac{2I}{I + mR^2}\right)v_- + \frac{2IR}{I + mR^2}\omega_- \\ \omega_+ &= \frac{2mR}{I + mR^2}v_- - \left(1 + \frac{2I}{I + mR^2}\right)\omega_-. \end{aligned} \tag{2.1.5}$$

For the two dimensional disk of uniform density, the moment of inertia is $I = \frac{1}{2}mR^2$, yielding

(for the upper collision)

$$\begin{aligned} v_+ =& \frac{1}{3}v_- + \frac{2}{3}R\omega_- \\ \omega_+ =& \frac{4}{3R}v_- - \frac{1}{3}\omega_-. \end{aligned} \tag{2.1.6}$$

In applying the model subsequently we divide the rotational coordinate by a factor of $\frac{R}{\sqrt{2}}$ so that the energy will be normalized. The collision transformation becomes

$$\begin{aligned} v_+ =& \frac{1}{3}v_- + \frac{2\sqrt{2}}{3}\omega_- \\ \omega_+ =& \frac{2\sqrt{2}}{3}v_- - \frac{1}{3}\omega_-. \end{aligned} \tag{2.1.7}$$

Adjusting the signs yields a similar transformation for collisions on the floor. More generally, though, we wish to consider the transformation according to a frame with tangential coordinate x_1, upward normal x_2, and x_0 representing the rotational direction. Then we may describe the collision as a transformation of the velocity $v = (\dot{x_0}, \dot{x_1}, \dot{x_2})^\dagger$, where '†' indicates matrix transpose. More specifically, the transformation matrix $T \in O(3)$ is

$$T = \begin{pmatrix} -\frac{1}{3} & \frac{2\sqrt{2}}{3} & 0 \\ \frac{2\sqrt{2}}{3} & \frac{1}{3} & 0 \\ 0 & 0 & -1 \end{pmatrix}. \tag{2.1.8}$$

This transformation will also arise, with greater context, in Chapter 3.

Considering the corresponding vector equations, Tokieda [28] uses this derivation to find the dimension 3 model, noting that there is a "dull" solution and a second solution "worthy of a superball." For the three dimensional ball, the moment of inertia is $I = \frac{2}{5}MR^2$, yielding the vector equations

$$\begin{aligned} v_+ =& \frac{3}{7}v_- + \frac{4}{7}R\omega_- \\ \omega_+ =& \frac{10}{7R}v_- - \frac{3}{7}\omega_- \end{aligned} \tag{2.1.9}$$

which normalizes to

$$\begin{aligned} v_+ &= \frac{3}{7}v_- + \frac{2\sqrt{2}}{3}\omega_- \\ \omega_+ &= \frac{2\sqrt{2}}{3}v_- - \frac{1}{3}\omega_-. \end{aligned} \tag{2.1.10}$$

This will correspond to the rough rank two solution using the derivation in the next chapter. It is important to note, however, that the vector equations for the three dimensional case actually yield a range of solutions based on the rough rank 1 cases. Tokieda's model, like the other referenced physical models, implicitly exclude the intermediate cases with an assumption that the ball is "perfectly rough" [16].

Cross in particular conducts physical experiments tracking the reflection angles of tennis balls off surfaces of varying friction levels. As one might expect, the resulting collisions yield actual angles between the no-slip and specular angles, falling along a spectrum according to the friction. For low impact angles, complications due to rolling occur [14]. While the case of rolling rigid body collisions is touched upon in the next chapter, the application is beyond our current scope.

2.2 Examples

We now turn to several specific examples, beginning with the equilateral triangle. We will prove the assertion of periodicity from Chapter 1. (Recall Figure 1.3.)

Example 3. *The equilateral triangle billiard*

By Proposition 28, which will be proved in the Chapter 4, the wedge angle $\theta = \frac{\pi}{3}$ yields period four (or degenerate period two) orbits. As a result, the equilateral triangle has unique dynamics. Immediately, any orbit which starts at one side and fails to contact both of the remaining sides by the third collision will be periodic. It turns out, in fact, that *all* orbits are periodic.

Proposition 1. All nondegenerate orbits of the no-slip equilateral triangle billiard are periodic, with period two, three, four, or six.

Proof. First, we show that the velocities are periodic with the given possible values. Labeling the sides, consider the possible combinations keeping track only of which side the collision occurs on and not the exact location, looking at equivalence classes up to relabeling sides. Each possible orbit can be written as a sequence of ones, twos, and threes, without loss of generality beginning with 12.

Certain combinations are dynamically impossible: we can immediately rule out any sequence with repeated numbers, and 1213232 can be ruled out because any 4−cycle between two edges will repeat. Other combinations, like 12131 can be ruled out by comparing the dynamics with the wedge dynamics when side 3 is removed. The possible parameters when leaving side 1 for the second time are such that it must, in the wedge, return to side two, parameters which also ensure the orbit would next hit side 2 if instead it collides with wall 3. Figure 3 shows the combinatorial possibilities with dotted lines indicating equivalent forms and dynamically impossible combinations struck through.

Besides the four cycle 1212, the only possibilities may be expressed 1213231 and 1231231. Fixing an orientation for the triangle, the transformation of the velocity vector after a collision, relative to the new frame, will either be $S_1 = TR'_{\frac{\pi}{3}}$ or $S_2 = TR'_{-\frac{\pi}{3}}$. But the combined transformations for the two possibilities after the sixth collision are (in the given representation) $S_1 S_2^3 S_1^2 = I$ or $S_1^6 = I$, and the velocity is preserved.

It remains to show that the orbits are actually periodic, that is, that the orbits return to the same boundary point and are not merely parallel trajectories. Suppose that the orbit arrives back on the starting side but not at the same point, and consider subsequent returns after multiples of six collisions. The second set of trajectories will be parallel to the first all the way around and accordingly the sequence of walls will not change. Now consider the finite partition of the original wall by degenerate points where an orbit in the fixed direction

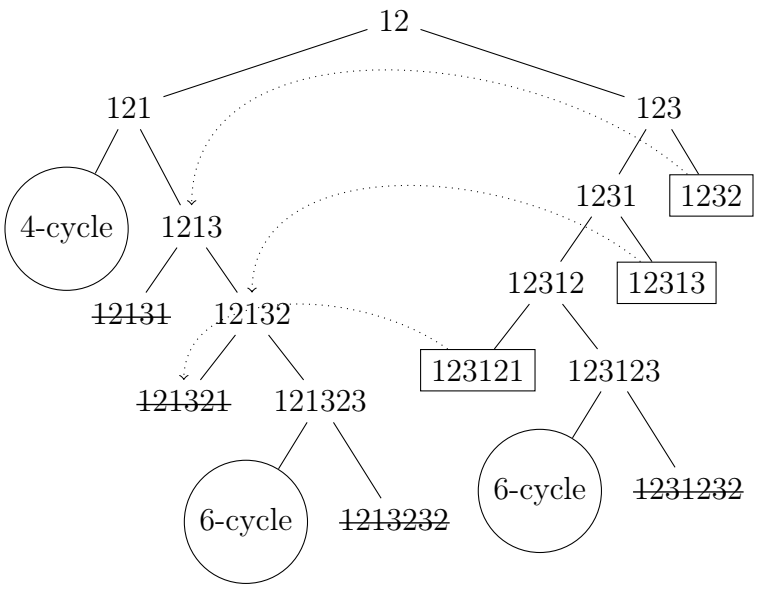

Figure 2.1: Combinatorial possibilities for trajectories of the equilateral triangle. Strike through indicates combinations which as impossible with the no-slip collision map, while boxes indicate duplicates.

from that point eventually hits a vertex. All return points must be in the same interval, always at the same distance apart and with the order preserved. But this implies that the entire interval is fixed, and the orbit must return to the starting point. □

Currently, the equilateral triangle is the only known example of a closed billiard we know of which is completely periodic in this sense. Good candidates for other examples are polygons with all periodic angles, in the sense of yielding the periodic wedge orbits as described in Theorem 3. Such polygons might turn out to be the no-slip analog of standard rational billiards, an active field of inquiry. (See, for example, Chapter 11 in [25] by Smillie or Chapter 13 in [4] by Masur and Tabachnikov.) However, even identifying "rational" polygons in this new sense is nontrivial, beyond the equilateral triangle.

We wish to illustrate now boundary conditions for which the map $q \mapsto \mathcal{C}_q$ varies in a nontrivial way or is chosen randomly. For the next two examples, let the billiard system consist of a disc moving in an infinite strip bounded by two parallel lines, as in Example 1,

but with an alternative method of choosing the collision model as designated.

Example 4. *The random half smooth, half no-slip disk*

First, we revisit the no-slip strip but with the requirement that a point on the boundary of the disc is chosen to be rough or smooth randomly with equal probability. This is thus an example of a random boundary condition. The longitudinal motion now corresponds to a random walk, (see 2.2) for which it is possible to prove a diffusion (Brownian motion) limit. (Under the assumption of unit velocity, if the ratio of the rotational to vertical components is $\sqrt{2}$, the system precisely corresponds to a unit random walk.)

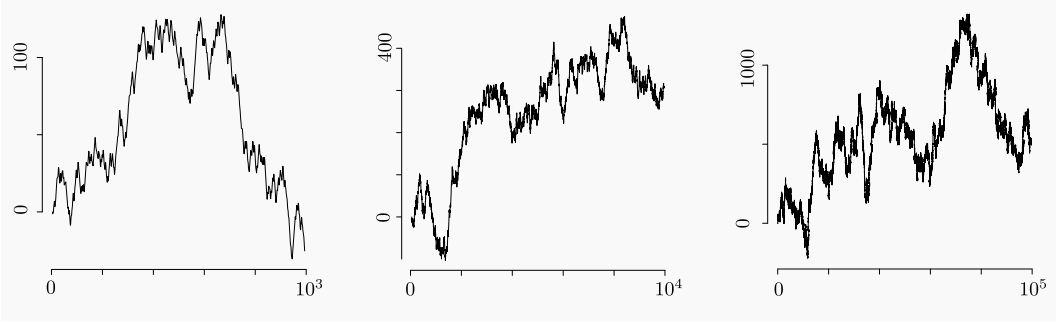

Figure 2.2: Disc between parallel plates in dimension 2. The roughness rank is now random, either 0 or 1 with equal probabilities.

In the deterministic case, the behavior is quite different.

Example 5. *The deterministic half smooth, half rough disk*

Suppose that one hemisphere of the boundary of the disc is rough, resulting in no-slip collisions, and the other is smooth, yielding specular collisions. In Figure 2.3 we show graphs of the position of the (center of) the disc along the longitudinal axis of the table as a function of the collision step. The time between two consecutive collisions is easily shown to be constant, so the step number is proportional to time. The three graphs describe the same trajectory at different time scales, as indicated in the legend of the figure.

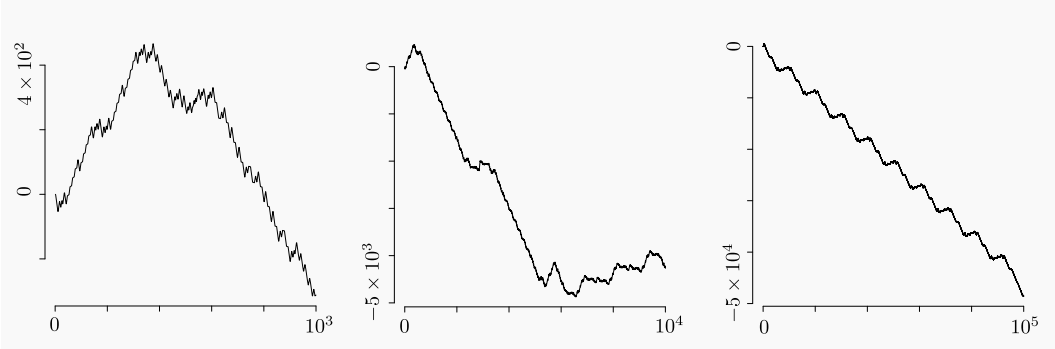

Figure 2.3: A single orbit of the motion of a disc between parallel plates in dimension 2. Half of the boundary circle is rough and the other half is smooth. The horizontal axis indicates the step number, taken as a proxy for time. The vertical axis gives the distance of the center of mass along the length of the 2-dimensional channel.

It is interesting to observe the apparent long range quasi-periodic behavior of trajectories.

We consider now a few examples in dimension 3. In all cases, a ball of uniform mass distribution moves between two parallel infinite plates. In dimension three, the roughness rank can be 0, 1, or 2. Standard specular reflection has roughness rank 0; we explore examples of roughness rank 1 and 2.

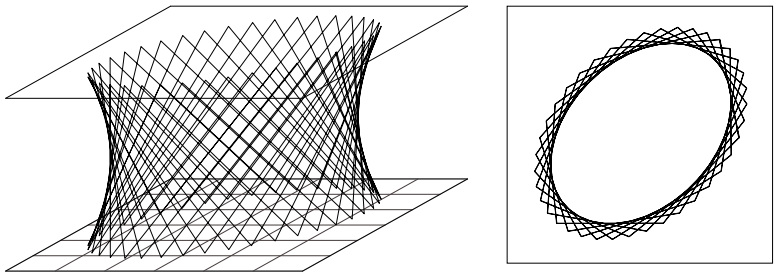

Figure 2.4: Typical segment of trajectory in position space and its horizontal projection for the motion of the center of mass of a ball bouncing between two parallel plates in dimension 3 with roughness rank equal to 2. Note that orbits are bounded, as will be shown in Chapter 4.

Example 6. *Three spacial dimensions, rough rank two*

Suppose the rough rank is two for both plates, that is, suppose we have the three dimensional analog of the no-slip strip. Simulations suggest that the trajectories are bounded

(Figure 2.4), and indeed we will prove this is the case.

The figure on the right shows the projection of the trajectory to the coordinate plane parallel to the plates.

Example 7. *Three spacial dimensions, rough rank one*

When the roughness rank is one, the set of collision maps comprise a one-dimensional family. Figure 2.5 shows some combinations of boundary conditions. One clearly notices that whenever the roughness rank in at least one plate is not maximal, trajectories are no longer bounded. The boundary conditions for the systems of Figure 2.5 are of the following types: one plate has roughness rank 2 and the other has roughness rank 1 with constant rough direction (that is, constant map $\mathcal{T}$ in Equation 3.1.2); and both plates have roughness rank 1, with random rough direction for the bottom plate and either constant or independent random rough direction for the top plate. Specifically, we choose the directions given by angles $0, \pi/3, 2\pi/3$ with equal probabilities. The legend of the figure shows which trajectory corresponds to which condition.

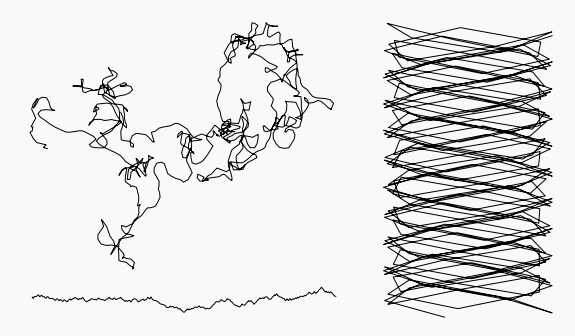

Figure 2.5: Horizontal projection of motion of the center of mass for a ball bouncing between two parallel plates in dimension 3. Top left: roughness rank 1 for both top and bottom plates and random (and independent) roughness directions; bottom left: top and bottom roughness rank 1, but now roughness direction is constant for the top and random for the bottom plate; right: roughness rank is 2 for top plate and 1 for bottom plate, with constant roughness direction. Making the the rough direction random for the bottom plate gives a trajectory that does not look significantly different than the one on the top left.

Chapter 3

A differential geometric framework for rigid body collisions in $\mathbb{R}^n$

This chapter gives the details of the larger theory of rigid body collisions described in [13]. It is joint work with Renato Feres and Will Ward. It is of interest here specifically because the no-slip model of Chapter 2 can be better understood as a special case, but is also of intrinsic interest and might serve as a framework for exploring other collision models.

The classical theory of collisions of rigid bodies provides a very natural setting in which to explore the geometry and dynamics of mechanical systems on configuration manifolds with boundary. From this geometric perspective, the response of the system to collisions between its rigid moving parts is specified by assigning appropriate boundary conditions that tell how a trajectory should be continued once it reaches the boundary. For example, in the theory of billiard dynamical systems, a topic that may be defined very broadly as the study of Hamiltonian (more typically, geodesic flow) systems on Riemannian manifolds with boundary, one typically assumes that trajectories reflect off the boundary specularly—the simplest form of impact response compatible with the basic laws of mechanics such as energy conservation and time reversibility. Billiard systems with more general boundary conditions have to our knowledge been investigated very rarely. Two pertinent examples are [5], discussed in Chapters 1 and 2, and [29], which will be discussed in Chapter 5. Both works are restricted to 2-dimensional billiards. (There is, of course, an extensive literature in engineering and applied physics about less idealized systems governed by impact interactions, including a few such as those discussed in the last chapter that touch upon rough collisions, but this

literature is not concerned with the differential geometric issues that are the main focus here.)

Our first goal is to classify boundary conditions for systems defined by two unconstrained rigid bodies in $\mathbb{R}^n$, $n \geq 1$, under standard physical assumptions of energy conservation, linear and angular momentum conservation, time reversibility, linearity of response, and another condition to be defined shortly that extends momentum conservation and is typically made implicitly in textbooks. Collisions satisfying all of these properties will be called *strict*. They are formally represented by linear maps $\mathcal{C}_q : T_qM \to T_qM$, where M is the configuration manifold equipped with the kinetic energy Riemannian metric, q is a boundary configuration, and the tangent space T_qM is the space of (pre- and post-) collision states. A boundary condition for the system then consists of the (differentiable, measurable, random, etc.) assignment of a strict collision map $\mathcal{C}_q$ to each boundary configuration $q \in \partial M$.

In dimensions greater than 1, the collision map is not uniquely determined by the conditions of strict collision. It is well-known that the nature of the contact between the colliding rigid bodies also needs to be specified. The standard case in which $\mathcal{C}_q$ is specular reflection corresponds to bodies having physically smooth surfaces.

Towards this classification we identify a family of subbundles of $T(\partial M)$ arising naturally under the assumed physical laws and discuss some relationships among them. From these relationships we derive sufficient conditions for the non-standard billiard system to leave invariant the natural volume measure on a constant energy manifold (derived from the canonical symplectic form). The invariance of this *billiard measure* makes it possible to bring the tools of ergodic theory (see [9, 27]) to the study of non-standard billiard systems; we will return to this in Chapter 5.

3.1 Statements of the main results

3.1.1 Notation, terminology, and standing assumptions

For the next sections we consider the unconstrained motion of two rigid bodies, represented by the sets $B_1, B_2 \subset \mathbb{R}^n$. We call these sets the *bodies in reference configuration*. Let $G = SE(n)$ denote the Euclidean group of orientation preserving isometries of $\mathbb{R}^n$ equipped with the standard inner product. The bodies are assumed to be connected n-dimensional submanifolds of $\mathbb{R}^n$ with smooth boundaries. An *interior configuration* of the system consists of a pair $(g_1, g_2) \in G \times G$ such that $g_1(B_1)$ and $g_2(B_2)$ are disjoint sets. The closure of the set of interior configurations, denoted M, will be called the *configuration space* of the system, and its boundary is the set of *contact* (or *collision*) configurations. M has dimension $2\dim(G)$ and the nature of the boundary ∂M will depend on geometric assumptions about the B_j. We will soon state a sufficient, and fairly general for our needs, condition on the ∂B_j for M to be a submanifold of $G \times G$ with smooth boundary.

The following notations are used fairly consistently throughout. Points in B_j are denoted b, b_j. Elements of the Euclidean group, which is the semidirect product $G = SO(n) \ltimes \mathbb{R}^n$ of the groups of rotations and translations, are written as pairs (A, a), where $A \in SO(n)$ and $a \in \mathbb{R}^n$. Elements of the Lie algebra $\mathfrak{g} = \mathfrak{se}(n)$ are written (Z, z), possibly with subscripts or superscripts, where $Z \in \mathfrak{so}(n)$ and $z \in \mathbb{R}^n$. The outward-pointing unit normal vector to the boundary of B_j at $b \in \partial B_j$ is denoted $\nu_j(b)$. It is convenient to consider orthonormal frames σ at $b \in \partial B_j$ *adapted* to the bodies, in the following sense: $\sigma : \mathbb{R}^n \to T_b\mathbb{R}^n \cong \mathbb{R}^n$ is an element of $SO(n)$ such that $\sigma e_n = -(-1)^j \nu_j(b)$, where $e_n = (0, \ldots, 0, 1)^\dagger$ is the last element of the standard basis $\{e_1, \ldots, e_n\}$ of $\mathbb{R}^n$ and '$\dagger$' indicates matrix transpose.

Because M is a submanifold of $G \times G$, each tangent space can be canonically identified with $T_{(e,e)}(G \times G) \cong \mathfrak{g} \oplus \mathfrak{g}$ by left-translation. Thus *states* of the system, defined as elements of TM, may be canonically identified with tuples $(A_1, a_1, A_2, a_2, Z_1, z_1, Z_2, z_2)$. We sometimes indicate

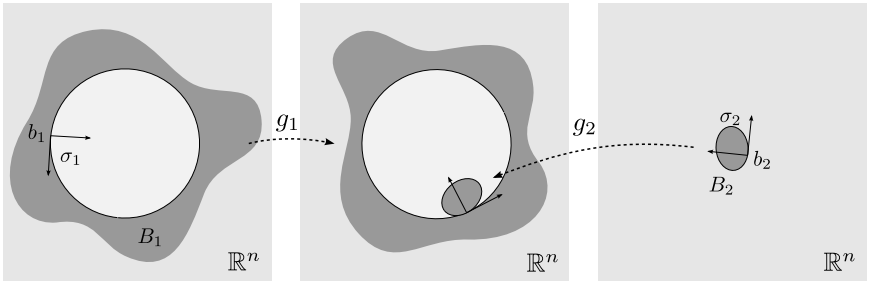

Figure 3.1: On the left and right are the bodies B_1 and B_2 in their reference configuration in $\mathbb{R}^n$. A configuration of the system of rigid bodies is given by a pair (g_1, g_2) of elements of the Euclidean group G. A boundary configuration can be parametrized by the tuple $(b_1, \sigma_1, b_2, \sigma_2)$ where $b_j \in \partial B_j$ such that $g_1(b_1) = g_2(b_2)$ and σ_j is an orthonormal frame at b_j as will be explained in the text.

the state by (q, ξ), where $q = (g_1, g_2)$ and $\xi = (Z_1, z_1, Z_2, z_2) \in \mathfrak{g} \times \mathfrak{g}$. The position of material point $b \in B_j$ in the given state is then $g_j(b) = A_j b + a_j$ and its velocity is $V(b) = A_j(Z_j b + z_j)$. The boundary configuration (g_1, g_2) can also be parametrized, up to an overall rigid motion of the two bodies keeping their positions relative to each other unchanged, by $(b_1, \sigma_1, b_2, \sigma_2)$, where b_j is in the boundary of B_j and σ_j is an adapted frame such that $g_1(b_1) = g_2(b_2)$ and $A_1\sigma_1 = A_2\sigma_2$. The tuple $(b_1, \sigma_1 h, b_2, \sigma_2 h)$ corresponds to the same contact configuration, for all $h \in H$, where

$$H := SO(n-1) = \{A \in SO(n) : Ae_n = e_n\}.$$

These notions are illustrated in Figure 3.1.

Let Π_b denote the orthogonal projection from $\mathbb{R}^n$ to the tangent space to the boundary of B_j at a boundary point b. The orthogonal projection to $\mathbb{R}^{n-1} = e_n^\perp$ will be denoted Π. The *shape operator* of the boundary of B_j at the point b is the linear map defined by

$$S_b : \nu_j(b)^\perp \to \nu_j(b)^\perp, \quad S_b v = -D_v \nu_j$$

where D_v is directional derivative in $\mathbb{R}^n$. We say that $S : \mathbb{R}^{n-1} \to \mathbb{R}^{n-1}$ is the shape operator S_b in the adapted frame σ at b if $\sigma S = S_b \sigma$. The notation $\mathrm{Ad}_\sigma S = \sigma S \sigma^{-1}$ will be used often

to indicate conjugation.

So as not to get distracted by regularity issues, we assume that the configuration manifold M has smooth boundary and that each boundary configuration corresponds to the bodies being in contact at a single common point. Proposition 2, which is a special case of Proposition 19, gives a sufficient condition for M to be nice in this respect. The hypotheses of the proposition will be assumed to hold subsequently.

Proposition 2. Suppose that S_1+S_2 is nonsingular for every relative configuration $(b_1, \sigma_1, b_2, \sigma_2)$ of the rigid bodies, where $S_j : \mathbb{R}^{n-1} \to \mathbb{R}^{n-1}$ is the shape operator of the boundary of B_j at b_j in the adapted frame σ_j. Then M is a smooth manifold of dimension $2\,dim G$ with smooth boundary, and each boundary point $q = (g_1, g_2)$ represents a configuration with a unique point of contact. Moreover the map that associates to $q \in \partial M$ the uniquely determined pair $(b_1, b_2) \in \partial B_1 \times \partial B_2$ such that $g_1(b_1) = g_2(b_2)$ is smooth.

We call the b_1, b_2 associated to $q \in \partial M$ under the condition of Proposition 2 the *contact points* (in the reference configuration) associated to boundary point q.

3.1.2 The kinematic bundles

If $a, b \in \mathbb{R}^n$, let $a \wedge b \in \mathfrak{so}(n)$ be the n-by-n matrix such that $(a \wedge b)_{ij} = a_j b_i - a_i b_j$. If a, b are orthogonal unit vectors, $a \wedge b$ is the infinitesimal generator of the one-parameter group in $SO(n)$ that rotates the plane spanned by a and b and fixes pointwise the orthogonal complement of that plane. The boundary state of the two-body system consists of the boundary configuration $q = (g_1, g_2) \in G \times G$ and velocities $\xi = (Z_1, z_1, Z_2, z_1) \in \mathfrak{g} \times \mathfrak{g}$.

We now define the *kinematic bundles*. Given $q = (g_1, g_2) \in \partial M$, with $g_j = (A_j, a_j)$ and associated contact points b_1, b_2, consider the following linear relations on the $\xi \in T_q(\partial M)$, where $N_j = \partial B_j$ and $\nu_j = \nu_j(b_j)$:

$$R_1 : \ \nu_1 \cdot (Z_1 b_1 + z_1) = \nu_2 \cdot (Z_2 b_2 + z_2)$$

$R_2:\ A_1(Z_1 b_1 + z_1) = A_2(Z_2 b_2 + z_2)$

$R_3:\ \mathrm{Ad}_{A_j} Z_j = W + \nu_j \wedge w_j$ for $W \in \mathfrak{so}(n)$ and $w_j \in T_{b_j} N_j$, $j = 1, 2$

$R_4:\ \mathrm{Ad}_{A_1} Z_1 = \mathrm{Ad}_{A_2} Z_2$.

As already noted, $A_j(Z_j b_j + z_j)$ is the velocity of the contact point b_j in the given state. Observe that relation R_2 implies R_1. It will be shown later that

$$T_q(\partial M) \cong \{\xi \in \mathfrak{g} \times \mathfrak{g} : R_1\}.$$

The physical interpretation of the kinematic bundles is as follows. A state satisfying R_1 has the property that the contact points have zero relative velocity in the normal direction to the plane of contact. Relation R_2 is satisfied exactly when the contact points are not moving at all relative to each other at the moment of contact. This means that the contact points do not slip past each other. Relation R_3 describes a state in which the tangent spaces to the bodies at the point of contact do not experience a relative rotation (on that tangent space). Thus it is a condition of *non-twisting*. Together R_2 and R_3 describe a state in which the bodies are rolling on each other.

Definition 2 (Kinematic bundles)**.** Let $\mathfrak{S}$, $\mathfrak{R}$, and $\mathfrak{S}$ be the vector subbundles of $T(\partial M)$ defined by

$$\mathfrak{S}_q \cong \{\xi \in \mathfrak{g} \times \mathfrak{g} : R_2\}$$

$$\mathfrak{R}_q \cong \{\xi \in \mathfrak{g} \times \mathfrak{g} : R_2, R_3\}$$

$$\mathfrak{D}_q \cong \{\xi \in \mathfrak{g} \times \mathfrak{g} : R_2, R_3, R_4\}.$$

Note that $\mathfrak{D} \subset \mathfrak{R} \subset \mathfrak{S}$. We refer to $\mathfrak{S}$ as the *non-slipping* subbundle, $\mathfrak{R}$ the *rolling* subbundle, and $\mathfrak{D}$ the *diagonal* subbundle.

It will be shown that the diagonal subbundle $\mathfrak{D}$ is the tangent bundle to the orbits of the action of G on M by left translations: $g(g_1, g_2) \coloneqq (gg_1, gg_2)$. Later in we give a different definition of these subbundles, Definition 10, that makes their physical interpretation more clear. Then what is stated above as a definition is derived in Section 3.4.2.

3.1.3 The kinetic energy metric and the impulse subbundle

Suppose now that the bodies B_j are assigned mass distributions represented by finite positive measures μ_j supported on B_j. Let $m_j \coloneqq \mu_j(B_j)$ be the *mass* of B_j. We may assume without loss of generality that μ_j has zero first moment: $\int_{B_j} b\, d\mu_j(b) = 0$. This is to say that B_j has *center of mass* at the origin of $\mathbb{R}^n$. The matrix of second moments of μ_j is $L_j = (l_{rs})$, with entries

$$l_{rs} = \frac{1}{m_j} \int_{B_j} b_r b_s \, d\mu_j(b).$$

We call L_j the *inertia matrix* of body B_j. This matrix induces a map $\mathcal{L}_j$ on $\mathfrak{so}(n)$ that associates to $Z \in \mathfrak{so}(n)$ the matrix $\mathcal{L}_j(Z) = L_j Z + Z L_j \in \mathfrak{so}(n)$.

Definition 3 (Kinetic energy Riemannian metric)**.** Given $q \in M$ and $u, v \in T_q M$, define the symmetric non-negative form on $T_q M$ by

$$\langle u, v \rangle_q = \sum_j m_j \left[\frac{1}{2} Tr\left(\mathcal{L}_j(Z_j^u) Z_j^{v\dagger} \right) + z_j^u \cdot z_j^v \right]$$

where $(Z_1^u, z_1^u, Z_2^u, z_2^u)$ and $(Z_1^v, z_1^v, Z_2^v, z_2^v)$ are the translates to $\mathfrak{g} \times \mathfrak{g}$ of u, v. When the above bilinear form is positive definite we call it the *kinetic energy Riemannian metric* on M. Denoting by $\| \cdot \|_q$ the corresponding norm at q, we call $\frac{1}{2}\|v\|_q^2$ the *kinetic energy* associate to state (q, v).

The kinetic energy function given in Definition 3 is easily shown (as indicated later) to come from integration with respect to the mass distribution measures of (one-half of) the

Euclidean square norm of the velocity of material point b over the disjoint union of B_1 and B_2. Thus Definition 3 agrees with the standard textbook definition of kinetic energy. It is also clear that the metric is invariant under the left-action of G on M. Note that the boundary of M is a G-invariant set.

For each $u \in \mathfrak{g}$ we define vector field $q \mapsto \tilde{u}_q \in T_qM$, $q = (g_1, g_2)$, by

$$\tilde{u}_q := \left.\frac{d}{dt}\right|_{t=0} e^{tu} q.$$

We call $u \mapsto \tilde{u}$ the *infinitesimal action* derived from the left G-action on M and $\tilde{u}$ the *vector field associated* to $u \in \mathfrak{g}$.

Definition 4 (Momentum map)**.** The map $\mathcal{P}^{\mathfrak{g}} : TM \to \mathfrak{g}^*$ defined by

$$\mathcal{P}^{\mathfrak{g}}(q, \dot{q})(u) = \langle \dot{q}, \tilde{u}_q \rangle_q$$

is called the *momentum map* associated to the G-action on M.

The most straightforward way of introducing dynamics into the system is through Newton's second law. There are several equivalent forms of it as we note later. The following is particularly convenient for our needs. We first define a *force field* (possibly time dependent) as a bundle map $F : TM \to T^*M$. Given a state $(q, \dot{q})$, $q = (g_1, g_2)$, each component F_j of F can be pulled-back to $\mathfrak{g}^*$ using right-translation R_{g_j}, so it makes sense to write

$$\frac{d}{dt}\mathcal{P}^{\mathfrak{g}}_j(q, \dot{q}) = R^*_{g_j} F_j. \tag{3.1.1}$$

This is Newton's second law written as a differential equation on the co-Lie algebra of G. Other useful forms are mentioned later. One of them is indicated in the next proposition, in which we use the notation $F^{\#}$ for the dual of F with respect to the left-invariant Riemannian

metric and write

$$(Y_j(t,q,\dot{q}), y_j(t,q,\dot{q})) := (dL_{g_j})_e^{-1} F^{\#}(t,q,\dot{q}) \in \mathfrak{g}.$$

Here we are using the differential of the left-translation map L_{g_j}.

Proposition 3. The equation $\frac{d}{dt}\mathcal{P}_j^{\mathfrak{g}}(q,\dot{q}) = R_{g_j}^* F_j$ is equivalent to

$$m_j\left(\mathcal{L}_j\dot{Z}_j - [\mathcal{L}_j Z_j, Z_j]\right) = \mathcal{L}_j(Y_j)$$
$$m_j\dot{v}_c = A_j y_j$$

where $g_j = (A_j, a_j)$ and $v_c = A_j z_j$ is the velocity of the center of mass of body B_j.

We assume that F results from the integrated effect of forces acting on the individual material points. That is, we assume that there exists a $\mathbb{R}^n$-valued measure φ_j on B_j parametrized by TM from which F is obtained by integration:

$$F(q,v)(u) = \int_{B_j} V_u(b) \cdot d\varphi_{j,q,v}(b)$$

for all $u \in T_qM$, where $V_u(b)$ is the velocity of the material point b in the state (q,u). Of special interest for us are the forces involved in the collision process. These *impulsive* forces are characterized by being very intense and of very short duration, applied on a single point—the point of contact in each body.

That the forces act on each body only at the point of contact greatly restricts the right-hand side of the equation of motion in Proposition 3. This is indicated in the next proposition.

Proposition 4. We suppose that the force field F_j acting on body B_j is such that the force distribution measure φ is singular, concentrated at the point b_j. Then the equations of motion

of Proposition 3 reduce to

$$m_j\left(\mathcal{L}_j\dot{Z}_j - [\mathcal{L}_j Z_j, Z_j]\right) = b_j \wedge y_j$$
$$m_j\dot{v}_c = A_j y_j$$

For ideal impulsive forces (of infinite intensity and infinitesimal duration), momentum should change discontinuously. Integrating Equation 3.1.1 over a very short time interval $[t^-, t^+]$ around t produces a nearly discontinuous change in momentum while keeping the configuration essentially unchanged. We have informally

$$\mathcal{P}_j^{\mathfrak{g}}(q,\dot{q}_+) - \mathcal{P}_j^{\mathfrak{g}}(q,\dot{q}_-) = \int_{t_-}^{t_+} R_{g_j}^* F_j\, ds = \text{ Impulse at } t.$$

It is not necessary for our needs to make more precise the limit process suggested by this expression. From it we obtain the form of the change in momentum after impact, which is given in the next proposition. Let $q = (g_1, g_2) \in \partial M$ be a collision configuration and denote by

$$(Z_1^{\pm}, z_1^{\pm}, Z_2^{\pm}, z_2^{\pm}) \in T_q M$$

the post- (+) and pre- (−) collision velocities of the two rigid bodies.

Proposition 5 (Velocity change due to impulse at contact point)**.** Given pre-collision velocity $(Z_1^-, z_1^-, Z_2^-, z_2^-)$ there exist $u_1, u_2 \in \mathbb{R}^n$ such that

$$z_j^+ = z_j^- + u_j$$
$$Z_j^+ = Z_j^- + \mathcal{L}_j^{-1}(b_j \wedge u_j).$$

Under conservation of linear momentum $m_1 A_1 u_1 + m_2 A_2 u_2 = 0$ holds.

The proof of the above proposition is given in Section 3.3.4. The assumption that impulsive

forces of one body on the other at the moment of impact are applied at the point of contact is a strong constraint. One can in principle conceive of force fields of relatively long range, acting throughout the bodies, that are briefly switched on at the moment of impact, then switched off as soon as the bodies lose contact. More realistically, the bodies could suffer a deformation around the region of impact, creating a small neighborhood of contact. Of course this goes beyond the rigid body model. Here it is assumed that these possibilities do not happen, and that any effect of one body on the other can only be transmitted through the single point of contact between them.

If L_j is non-negative definite of rank at least $n-1$, $\mathcal{L}_j$ is invertible. With this in mind, Proposition 5 suggests the following definition.

Definition 5 (Impulse subbundle)**.** The *impulse subbundle* of TM (over the base manifold ∂M) is defined so that its fiber at $q \in \partial M$ is the subspace

$$\mathfrak{C}_q = \left\{((\mathcal{L}_1^{-1}(b_1 \wedge u_1), u_1), (\mathcal{L}_2^{-1}(b_2 \wedge u_2), u_2)) : u_j \in \mathbb{R}^n, m_1 A_1 u_1 + m_2 A_2 u_2 = 0\right\}.$$

We have now the following vector subbundles of $i^*(TM)$, where $i : \partial M \to M$ is the inclusion map: $\mathfrak{D} \subset \mathfrak{R} \subset \mathfrak{S} \subset T(\partial M)$ and $\mathfrak{C}$. The latter subbundle is the only one that depends on the mass distributions.

Theorem 1. The impulse subspace $\mathfrak{C}_q$ is the orthogonal complement of the non-slipping subspace $\mathfrak{S}_q$ and contains the unit normal vector $\mathbb{n}_q$. Therefore,

$$T_q M = \mathfrak{S}_q \oplus (\mathfrak{C}_q \ominus \mathbb{R}\mathbb{n}_q) \oplus \mathbb{R}\mathbb{n}_q$$

is an orthogonal direct sum.

A physical interpretation of this orthogonal decomposition will be given shortly.

3.1.4 Collision maps

Let $\mathbb{n}_q$ be the unit normal vector to ∂M pointing into M at a boundary configuration q. Define the half-spaces

$$T_q^+ M := \{v \in T_q M : \langle v, \mathbb{n}_q \rangle \geq 0\} = -T_q^- M.$$

We call any $\mathcal{C}_q : T_q^- M \to T_q^+ M$ a *collision map* at q. By a *boundary condition* we mean the assignment of such a map $\mathcal{C}_q$ to each $q \in \partial M$. We only consider here *linear* collision maps; that is, $\mathcal{C}_q$ extends to a linear map on $T_q M$. We now restate Definition 1 more formally.

Definition 6 (Strict collision maps)**.** A collision map $\mathcal{C}_q$ at $q \in \partial M$ is *strict* if the following hold for all $u, v \in T_q M$:

1. *Conservation of energy*: $\langle \mathcal{C}_q v, \mathcal{C}_q u \rangle_q = \langle v, u \rangle_q$. That is, $\mathcal{C}_q$ is a linear isometry.
2. *Conservation of momentum*: $\mathcal{P}^{\mathfrak{g}}(q, \mathcal{C}_q v) = \mathcal{P}^{\mathfrak{g}}(q, v)$.
3. *Time reversibility*: $\mathcal{C}_q^2 = Id$ (a linear involution).
4. *Impulse at the point of contact*: $\mathcal{C}_q v - v \in \mathfrak{C}_q$.

Proposition 6. Condition 2 of Definition 6 is equivalent to assuming that $\mathcal{C}_q$ restricts to the identity map on $\mathfrak{D}_q$. Condition 4 is equivalent to $\mathcal{C}_q \mathfrak{C}_q = \mathfrak{C}_q$ and $(\mathcal{C}_q - Id)\mathfrak{C}_q^{\perp} = 0$.

Thus energy conservation and impulse at a single contact point are together equivalent to $\mathcal{C}_q$ being the identity on $\mathfrak{S}_q$. In this sense, condition (4) of Definition 6 can be regarded as generalizing momentum conservation as we note in Proposition 6. In fact, conservation of momentum amounts to $\mathcal{C}_q$ being the identity on $\mathcal{D}_q$, whereas 4 and Theorem 1 imply that $\mathcal{C}_q$ is the identity on the bigger subspace $\mathfrak{S}_q$. An intermediate condition is that $\mathcal{C}_q$ restricts to the identity on the rolling subspace $\mathfrak{R}_q$.

Corollary 1. Strict collision maps are the linear isometric involutions of T_qM, $q \in \partial M$, that restrict to the identity map on the non-slipping subspace $\mathfrak{S}_q$.

Collision maps have eigenvalues ± 1. The map $P_\pm := \frac{I \pm \mathcal{C}_q}{2}$ is the orthogonal projection to the eigenspace associated to eigenvalue ± 1.

Definition 7. The dimension of the eigenspace of $\mathcal{C}_q$ associated to eigenvalue -1 will be called the *roughness rank* of $\mathcal{C}_q$. The image of the orthogonal projection P_- (a subspace of $\mathfrak{C}_q$) will be called the *roughness subspace* at q.

The unit normal vector $\mathtt{n}_q$ is always contained in the impulse subspace $\mathfrak{C}_q$ and it must necessarily be in the -1-eigenspace of $\mathcal{C}_q$.

Corollary 2. Let n be the dimension of the ambient Euclidean space. Identifying

$$\mathfrak{C}_q \ominus \mathbb{R}\mathtt{n}_q \cong \mathbb{R}^{n-1},$$

the set of strict collision maps is the set of $C \in O(n-1)$ such that $C^2 = I$. Writing

$$\mathcal{J}_k := O(n-1)/(O(n-k-1) \times O(k)),$$

then the set of strict collision maps at any given boundary point is $\mathcal{J}_0 \cup \cdots \cup \mathcal{J}_{n-1}$. Moreover, $\dim \mathcal{J}_k = k(n-k-1)$ and k is the roughness rank at q.

We call $\mathcal{J}_k$ the *Grassmannian of rough subspaces* having roughness rank k. The classification theorem give in Chapter 1, which we restate here, follows.

Theorem 2. At each boundary point of the configuration manifold of the system of two rigid bodies, assuming the boundary is differentiable at that point, the set of all strict collision maps can be expressed as the disjoint union of orthogonal Grassmannian manifolds $Gr(k, n-1)$,

$k = 0, \ldots, n-1$, of all k dimensional planes in $\mathbb{R}^{n-1}$. In particular, when $n = 2$, the set of strict collision maps is a two-point set consisting of the specular reflection and the no-slip collision.

It is easy to compute the dimensions of the Grassmannians $\mathcal{J}_k$ for strict collision maps. They are given, up to dimension 5, by the following table:

$$\dim \mathcal{J}_k : \left[\begin{array}{c|ccccc} k & 0 & 1 & 2 & 3 & 4 \\ n & & & & & \\ \hline 1 & 0 & & & & \\ 2 & 0 & 0 & & & \\ 3 & 0 & 1 & 0 & & \\ 4 & 0 & 2 & 2 & 0 & \\ 5 & 0 & 3 & 4 & 3 & 0 \end{array}\right]$$

The table shows that in dimension 1 there is a unique strict collision map; in dimension 2 there are exactly 2 possibilities; and in dimension 3 there is one possibility of roughness rank 0 given by the standard reflection map, one possibility for maximal roughness rank 2, and a one-dimensional set of possibilities for roughness rank 1 parametrized by the lines through the origin in $\mathbb{R}^2$. For general n, the unique collision map of maximal roughness rank will be referred to as the *completely rough* reflection map.

3.1.5 Non-standard billiard systems

We have so far considered systems consisting of two unconstrained rigid bodies. The results can be extended to situations in which one body or both are subject to holonomic and non-holonomic constraints. Here we consider only the case in which body B_1 remains fixed in place whereas B_2 is unconstrained except for the condition that it cannot overlap with B_1. The term *billiard system* will refer to a system of this kind where B_2 is a ball with rotationally

symmetric mass distribution. The system will be called non-standard if the (strict) collision maps are not all specular reflection.

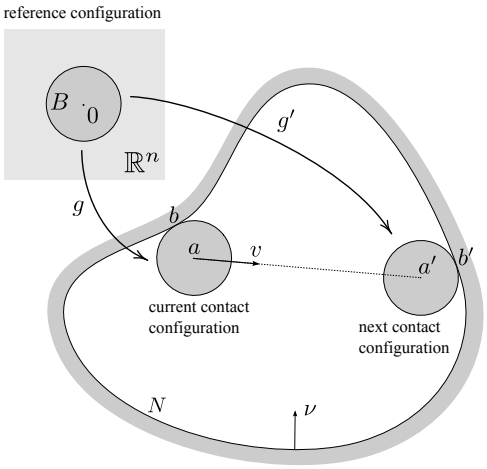

Figure 3.2: A billiard system. Body B_1 is kept fixed (the billiard table) and $B := B_2$ is a ball of rotationally symmetric mass distribution that can move freely in the complement of B_1. Given the contact point $b \in N$ and the post-collision velocity v of the center of mass of B, the point of contact of the next collision will be written $b' = T(b, v)$.

Let R denote the radius of $B := B_2$ and m its mass. Due to rotational symmetry of the mass distribution of B, the matrix of inertia L is scalar, that is, $L = \lambda I$. For example, a simple integral calculation shows that if B has uniform mass distribution, then $\lambda = \frac{R^2}{n+2}$. (Recall that we have defined L as the matrix of second moments of the mass distribution measure divided by the total mass, so m does not appear in λ.) The (smooth) boundary of B_1 will be denoted N and the unit normal vector field on N pointing into the region of free motion of B will be denoted ν. Trajectories of the billiard system are sequences of states: $(g_0, \xi_0), (g_1, \xi_1), \ldots$, where

$$(g_i, \xi_i) \in SE(n) \times \mathfrak{se}(n) \cong TSE(n), \quad g_i = (A_i, a_i), \quad \xi_i = (Z_j, z_j).$$

Here g_i is the contact configuration and ξ_i the post-collision velocities in the body frame

(reference configuration) at the ith collision.

To each contact state $(g, \xi) = (A, a, Z, z)$ is associated a unique contact point $b \in N$ and the post-collision velocity $v = Az$ of the center of mass. The center of mass of B in configuration g is a and the velocity of any given material point $b \in B$ is $V(b) := A(Zb + z)$. When it is necessary to distinguish points in N and in B we write $b \in N$ and $b_\circ \in B$. The unit normal vector to B at $b_\circ$ will be written $\nu_\circ(b_\circ) = b_\circ/R$. The point of contact at the next collision, which only depends on b and v, will be denoted $b' = T(b, v)$. See Figure 3.2.

One step of the billiard motion, $(g, \xi) \mapsto (g', \xi')$, amounts to the following operations.

1. From the current collision state (g, ξ) at time t one obtains the contact point $b \in N$ and velocity v of the center of mass a of B where $g = (A, a) = (A(t), a(t))$. It should be kept in mind that $\xi = (Z, z)$ describe post-collision velocities so $v = Az$ points into the region of free motion of the ball.

2. Obtain the contact point $b' = T(b, n) \in N$ and the time $t' = t + \tau$ of the next collision.

3. Obtain the next pre-collision state: (g', ξ^-) where $g' = (A', a') = (A(t + \tau), a(t + \tau))$, $\xi^- = (Z^-, z^-)$, and

$$A' = Ae^{\tau Z},\ a' = a + \tau Az,\ Z^- = Z,\ z^- = e^{-\tau Z} z.$$

   This is the free (geodesic) motion between collisions. Observe that $a' = b' + R\nu(b')$.

4. Let $b_\circ = (g')^{-1} b' \in \partial B$ be the contact point on the ball in the reference configuration at the next collision and denote by $\Pi_\circ, \Pi_\circ^\perp$ the orthogonal projections to the tangent space to ∂B at $b_\circ$ and to $\mathbb{R}\nu_\circ(b_\circ)$, respectively. Note that $A'\nu_\circ(b_\circ) = -\nu(b')$.

5. Finally, compute $\xi' = (Z', z')$ from (Z^-, z^-) using the choice of collision map. It will be

shown that

$$(Z', z') = \left(Z^- - \frac{\alpha}{2\lambda} b_\circ \wedge (I - \mathcal{T}) V^-, z^- - \alpha (I - \mathcal{T}) V^- - 2\Pi_\circ^\perp z^- \right), \tag{3.1.2}$$

where $\alpha := 1/(1 + R^2/2\lambda)$, $V^- = \Pi_\circ(Z^- b_\circ + z^-)$, and $\mathcal{T}$ is a linear involution on $T_{b_\circ}(\partial B)$ corresponding to a choice of collision map. For specular reflection $\mathcal{T} = I$ and for completely rough collisions $\mathcal{T} = -I$.

We assume the uniform mass distribution on the ball B so $\lambda = R^2/(n+2)$, where R is the radius of B. Let first $n = 2$. In this case the only non-standard collision map corresponds to $\mathcal{T} = -I$. Elements of the rotation group are parametrized by the angle of rotation θ and elements of the Lie algebra of $SO(2)$ are written as $\dot{\theta} J$, where J is the rotation matrix by $\pi/2$ in the counterclockwise direction. Together with the standard coordinates (x, y) we obtain coordinates (θ, x, y) on $SE(2)$. It will be convenient to make the coordinate change: $x_0 = R\theta/\sqrt{2}, x_1 = x, x_2 = y$. This yields coordinates $(x_0, x_1, x_2, \dot{x}_0, \dot{x}_1, \dot{x}_2)$ on the billiard state space. We also write $v_0 = \dot{x}_0$ and $v = (\dot{x}_1, \dot{x}_2)^\dagger$ for the velocity of the center of mass of the disc.

The choice of coordinates is made so that the kinetic energy Riemannian metric becomes, up to multiplicative constant, the standard Euclidean metric. Then it can be derived from Equation 3.1.2 that the post-collision velocities (v_0^+, v^+) after collision at point of contact $b \in N$ is the function of the pre-collision velocities (v_0^-, v^-) given by

$$\begin{aligned} v_0^+ &= -\frac{1}{3} v_0^- + \frac{2\sqrt{2}}{3} v \cdot (J\nu(b)) \\ v^+ &= \left[\frac{2\sqrt{2}}{3} v_0^- + \frac{1}{3} v^- \cdot (J\nu(b)) \right] J\nu(b) - v^- \cdot \nu(b)\nu(b). \end{aligned} \tag{3.1.3}$$

Thus the state updating equations for a 2-dimensional non-standard billiard system is as follows. If τ is the time of free flight between the two consecutive collisions and setting

$\mathbf{x} = (x_0, x_1, x_2)$, $\mathbf{v} = (v_0, v_1, v_2)$, then the billiard map giving the next state $(\mathbf{x}', \mathbf{v}')$ as a function of the present state $(\mathbf{x}, \mathbf{v}^-)$ is $(\mathbf{x}', \mathbf{v}') = (\mathbf{x} + \tau\mathbf{v}, \mathbf{v}^+)$ where $\mathbf{v}^+$ is related to $\mathbf{v}^-$ according to Equations 3.1.3. For a geometric interpretation of those equations, refer back to Figure 1.4. Note the role played by the angle β defined by $\cos\beta = 1/3$, $\sin\beta = 2\sqrt{2}/3$. In [5] it is observed that β is the dihedral angle of a regular tetrahedron. In the figures throughout we only indicate the position of the center of the disc; we draw a smaller table whose boundary is at a distance R from the boundary of the original table and we imagine the center of the ball as a point mass bouncing off the boundary of this smaller region.

3.2 The Euclidean group and its Lie algebra

The proofs of the main statements made above will be given after we establish some basic material. Despite the classical nature of the subject, standard textbook treatments of collisions of rigid bodies are not adequate for our needs while the more differential geometric texts in mechanics mostly do not treat this topic. Thus we find it necessary to develop the subject more or less from scratch. In this section we review general facts about the Lie theory and Riemannian geometry of the Euclidean group with left-invariant metrics that will be needed thoughout the remainder of this chapter.

3.2.1 Generalities

The *isometry group* of $(\mathbb{R}^n, \cdot)$, where '$\cdot$' indicates the standard inner product, is the Lie group of all the affine maps of the form $x \mapsto Ax + a$ for $A \in O(n)$ and $a \in \mathbb{R}^n$ under composition of maps. The closed subgroup of orientation preserving isometries, in which $A \in SO(n)$, is the *Euclidean group* in dimension n, denoted $SE(n)$. The latter is isomorphic to the semidirect

product $SO(n) \ltimes \mathbb{R}^n$ with multiplication operation

$$(A_2, a_2)(A_1, a_1) = (A_2A_1, A_2a_1 + a_2)$$

and inverse

$$(A, a)^{-1} = (A^{-1}, -A^{-1}a).$$

It is also isomorphic to a subgroup of the general linear group $GL(n+1, \mathbb{R})$ under the correspondence

$$(A, u) \in SO(n) \ltimes \mathbb{R}^n \mapsto \begin{pmatrix} A & u \\ 0 & 1 \end{pmatrix} \in GL(n+1, \mathbb{R}).$$

The Lie algebras of $SO(n)$ and $SE(n)$ will be denoted $\mathfrak{so}(n)$ and $\mathfrak{se}(n)$. The former consists of all the skew-symmetric matrices in the linear space $M(n, \mathbb{R})$ of $n \times n$ real matrices and $\mathfrak{se}(n)$, when $SE(n)$ is viewed as a subgroup of $GL(n+1, \mathbb{R})$, consists of the matrices

$$\begin{pmatrix} X & x \\ 0 & 0 \end{pmatrix} \in M(n+1, \mathbb{R})$$

where $X \in \mathfrak{so}(n)$ and x is any vector in $\mathbb{R}^n$. Indicating the matrix by the pair (X, x), the Lie bracket is written

$$[(X, x), (Y, y)] = (XY - YX, Xy - Yx).$$

One-parameter subgroups of $SE(n)$ have the form

$$\sigma(t) := \exp\left(t \begin{pmatrix} X & w \\ 0 & 0 \end{pmatrix}\right) = \begin{pmatrix} e^{tX} & \int_0^t e^{sX} w \, ds \\ 0 & 1 \end{pmatrix}.$$

It is useful to introduce the *wedge product*, the bilinear operation that associates to a pair

of vectors a, b in $\mathbb{R}^n$ the skew-symmetric matrix $a \wedge b \in \mathfrak{so}(n)$ whose (i,j)-entry is

$$(a \wedge b)_{ij} = a_j b_i - a_i b_j.$$

The following elementary properties of the wedge product will be used. The transpose of a matrix will be indicated by $U^\dagger$.

Proposition 7. Let a, b, u be (column) vectors in $\mathbb{R}^n$, $A \in SO(n)$ and $Z \in \mathfrak{so}(n)$. Then

1. $(a \wedge b)u = (a \cdot u)b - (b \cdot u)a$
2. $(a \wedge b)^\dagger = b \wedge a$
3. $A(a \wedge b)A^{-1} = (Aa) \wedge (Ab)$
4. $Tr((a \wedge b)Z^\dagger) = 2(Za) \cdot b$
5. $Tr((a \wedge b)(c \wedge d)^\dagger) = (a \cdot c)(b \cdot d)$
6. Let V be the span of orthogonal unit vectors $a, b \in \mathbb{R}^n$. Then $(a \wedge b)^2 = -I$ and

   $$R(\theta) := \exp(\theta a \wedge b) = (\cos\theta)I + (\sin\theta)a \wedge b \in SO(n).$$

   Thus $R(\theta)$ is the identity on $V^\perp$, and a rotation on V.
7. Let $e_n = (0, \ldots, 0, 1)^\dagger \in \mathbb{R}^n$ and $\Pi : \mathbb{R}^n \to \mathbb{R}^{n-1} = e_n^\perp$ the orthogonal projection. Then

   $$Z = \Pi Z \Pi + e_n \wedge (Ze_n)$$

   and $\Pi Z \Pi = 0$ iff there exists $z \in \mathbb{R}^{n-1}$ such that $Z = e_n \wedge z$.
8. For $a \in \mathbb{R}^3$ set $\omega(a)b := a \times b$—the cross-product by a on the left. Then $a \wedge b = \omega(a \times b)$ and $A\omega(a)A^{-1} = \omega(Aa)$.

9. If $n = 2$, then $a \wedge b = b \cdot (Ja)J$ where J is counterclockwise rotation by $\pi/2$.

Proof. All properties are proved by straightforward calculations. □

3.2.2 Left-invariant Riemannian metrics on $SE(n)$

Let $\langle \cdot, \cdot \rangle$ be a left-invariant Riemannian metric on the connected Lie group G with Lie algebra $\mathfrak{g}$. Let ∇ be the associated Levi-Civita connection. Define $B : \mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$ by

$$\langle B(u, v), w \rangle = \langle [v, w], u \rangle.$$

If X, Y, Z are left-invariant vector fields on G such that $X_e = u, Y_e = v, Z_e = w$, then from

$$2\langle \nabla_X Y, Z \rangle = -\langle [Y, Z], X \rangle - \langle [X, Z], Y \rangle + \langle [X, Y], Z \rangle \tag{3.2.1}$$

we obtain

$$(\nabla_X Y)_e = \frac{1}{2} \{ [u, v] - B(u, v) - B(v, u) \}. \tag{3.2.2}$$

A left-invariant vector field X is a *geodesic* vector field if and only if $0 = \nabla_X X = -B(X, X)$. It is not difficult to show that if the metric is bi-invariant then $B(u, u) = 0$ for all $u \in \mathfrak{g}$.

We adopt the notation: If $v \in T_g G$, then $g^{-1} v := \left(dL_{g^{-1}} \right)_g v \in T_e G = \mathfrak{g}$.

Proposition 8. Let $g(t)$ be any smooth curve in G and X a vector field along $g(t)$, not necessarily left-invariant. Define $z(t) := g(t)^{-1} \dot{g}(t)$ and $w(t) := g(t)^{-1} X_{g(t)}$. Then

$$g(t)^{-1} \left(\frac{\nabla X}{dt} \right)_{g(t)} = \dot{w} + \frac{1}{2} \left([z, w] - B(z, w) - B(w, z) \right).$$

In particular, $g(t)$ is a geodesic if and only if $\dot{z} = B(z, z)$.

Proof. Let $e_1, \ldots, e_n$ be a basis of $\mathfrak{g}$ and $E_1, \ldots, E_n$ the respective left-invariant vector fields

on G. We write $X = \sum f_j E_j$, $g'(t) = \sum h_j(t) E_j(g(t))$. Then, using Equation 3.2.2,

$$\frac{\nabla X}{dt} = \sum_{j,k} h_j'(t) \left[E_j f_k + \frac{1}{2} f_k \left([E_j, E_k] - B(E_j, E_k) - B(E_k, E_j) \right) \right]_{g(t)},$$

from which we obtain the desired expression after left-multiplying by $g(t)^{-1}$. □

Let the Lie algebra of G be $\mathfrak{g} = \mathfrak{s} \oplus \mathfrak{r}$, where $\mathfrak{r}$ is an ideal and $\mathfrak{s}$ is a Lie subalgebra. Let $\langle \cdot, \cdot \rangle$ be a left-invariant Riemannian metric on G and ∇ the corresponding Levi-Civita connection. We suppose that $\mathfrak{s}$ and $\mathfrak{r}$ are orthogonal subspaces.

Proposition 9. The following properties hold, where we indicate by the same letter elements of $\mathfrak{g}$ and the associated left-invariant vector fields on G. For $z, z_j \in \mathfrak{r}$, $Z, Z_j \in \mathfrak{s}$

1. $\nabla_{Z_1} Z_2$ and $B(Z_1, Z_2)$ lie in $\mathfrak{s}$
2. $\nabla_{z_1} z_2 \in \mathfrak{r}$. If $\mathfrak{r}$ is abelian, $\nabla_{z_1} z_2 = 0$ and $B(z_1, z_2) \in \mathfrak{s}$.
3. $\nabla_z Z = 0$ and $\nabla_Z z = [Z, z]$. Moreover $B(z, Z) \in \mathfrak{r}$ and $B(Z, z) = 0$.

Proof. All properties follow from the definition of B, Expression 3.2.1 for the Levi-Civita connection, and the assumption that the subalgebra $\mathfrak{s}$ and the ideal $\mathfrak{r}$ are orthogonal. □

Let S and R be the subgroups of G having lie algebras $\mathfrak{s}$ and $\mathfrak{r}$, respectively. Then G is the semi-direct product $G = S \ltimes R$, where R is a normal subgroup of G. We now assume that R is a vector subgroup, hence abelian, and that S is a compact subgroup acting on R by linear transformations, $S \subset GL(R)$, preserving an inner product $\langle \cdot, \cdot \rangle$ on R. That is, S is a subgroup of the orthogonal group $O(R, \langle \cdot, \cdot \rangle)$. Elements of G will be denoted (A, a) where $A \in S$ and $a \in R$. Indicating the action of S on R by Aa, the multiplication in G takes the form

$$(A_1, a_1)(A_2, a_2) = (A_1 A_2, A_1 a_2 + a_1).$$

Note that $\mathrm{Ad}_{(A,0)}(0,z) = (0,Az)$ and $\mathrm{ad}_{(Z,0)}(0,z) = (0,Zz)$, where $\langle Zz, z\rangle = 0$ since A acts on R by isometries.

Proposition 10. Under the just stated assumptions $B(z,z) = 0$ and $B(z,Z) = -Zz$. If $g(t)$ is a geodesic, writing $(Z(t), z(t)) = g(t)^{-1}\dot{g}(t)$ we have $\dot{Z} = B(Z,Z)$ and $\dot{z} = -Zz$

Proof. These simple remarks are consequences of the definition of B, the algebraic assumptions about the group, and Proposition 8. □

3.3 Newtonian mechanics of rigid bodies

We give here some alternative expressions of Newton's equation of motion. The approach, if not the notations, is essentially that of [2]. Other useful references are [3] and [21].

3.3.1 Momentum of a tangent vector and the momentum map

If M is a Riemannian manifold with metric $\langle\cdot,\cdot\rangle$ and $v \in T_qM$, we denote

$$\mathcal{P}(q,v) := \langle v, \cdot\rangle_q \in T_q^*M$$

and call this covector the *momentum* associated to the (velocity) vector v at (configuration) q. The pair (q,v) will be called a *state* of the system. We often indicate states by $(q,\dot{q})$, dotting the quantities of which time derivative is taken. The momentum map $\mathcal{P}^{\mathfrak{g}} : TM \to \mathfrak{g}^*$ was defined in Section 3.1.3.

If $M = G$ is endowed with a left-invariant Riemannian metric, the momentum map for the left-action of G is given by $\mathcal{P}^{\mathfrak{g}}(g,\dot{g})(u) = \langle \dot{g}, (dR_g)_e\, u\rangle_g$. Because the Riemannian metric is left-invariant,

$$\mathcal{P}^{\mathfrak{g}}(g,\dot{g})(u) = \langle g^{-1}\dot{g}, \mathrm{Ad}_{g^{-1}} u\rangle_e.$$

In this case we also write $\mathcal{P}^{\mathfrak{g}}(v) \coloneqq \langle v, \cdot \rangle_e \in \mathfrak{g}^*$ for $v \in \mathfrak{g}$. Then

$$\mathcal{P}^{\mathfrak{g}}(g, \dot{g}) = \mathrm{Ad}^*_{g^{-1}} \mathcal{P}^{\mathfrak{g}}(g^{-1}\dot{g})$$

where $\mathrm{Ad}^*_g \alpha = \alpha \circ \mathrm{Ad}_g$ and Ad_g is the differential of the map $L_g \circ R_{g^{-1}}$.

Proposition 11. Given a smooth curve $g(t)$ in G and setting $z(t) \coloneqq g(t)^{-1}\dot{g}(t)$, then

$$\frac{d}{dt}\mathcal{P}^{\mathfrak{g}}(g, \dot{g}) = Ad^*_{g^{-1}} \mathcal{P}^{\mathfrak{g}}(\dot{z} - B(z, z)).$$

In particular, $g(t)$ is a geodesic if and only if momentum $\mathcal{P}^{\mathfrak{g}}(g, \dot{g})$ is constant.

Proof. First note that

$$\frac{d}{dt}\mathrm{Ad}_{g^{-1}}(u) = -[z, \mathrm{Ad}_{g^{-1}} u].$$

It follows from the definitions that

$$\frac{d}{dt}\mathcal{P}^{\mathfrak{g}}(g, \dot{g})(u) = \frac{d}{dt}\langle z, \mathrm{Ad}_{g^{-1}} u \rangle = \langle \dot{z}, \mathrm{Ad}_{g^{-1}} u \rangle - \langle z, [z, \mathrm{Ad}_{g^{-1}}] \rangle = \langle \dot{z} - B(z, z), \mathrm{Ad}_{g^{-1}} u \rangle.$$

The expression on the far right is now $\mathcal{P}^{\mathfrak{g}}(\dot{z} - B(z, z)) \circ \mathrm{Ad}_{g^{-1}}$ evaluated at u. □

When $G = SO(n)$, define on $M(n, \mathbb{R})$ the bilinear form

$$\langle X, Y \rangle_0 \coloneqq \mathrm{Tr}(XY^\dagger).$$

Then $\langle \cdot, \cdot \rangle_0$ is an Ad_G-invariant non-degenerate positive bilinear form on $\mathfrak{so}(n)$ and the associated left-invariant Riemannian metric on G is bi-invariant. Thus for any left-invariant Riemannian metric $\langle \cdot, \cdot \rangle$ there must exist a linear map $\mathcal{L} : \mathfrak{g} \to \mathfrak{g}$, symmetric and positive definite with respect to $\langle \cdot, \cdot \rangle_0$, such that $\langle Z_1, Z_2 \rangle = \frac{1}{2}\langle \mathcal{L}(Z_1), Z_2 \rangle_0$. We are interested in such $\mathcal{L}$ that arises from a symmetric matrix $L \in M(n, \mathbb{R})$ according to the definition $\mathcal{L}(Z) \coloneqq ZL + LZ$,

in which case

$$\frac{1}{2}\mathrm{Tr}(\mathcal{L}(Z_1)Z_2^\dagger) = \mathrm{Tr}\left(Z_1 L Z_2^\dagger\right).$$

If $u_1, \ldots, u_n$ is a basis of $\mathbb{R}^n$ of eigenvectors of L, $Lu_i = \lambda_i u_i$, then the $u_i \wedge u_j$ comprise a basis of $\mathfrak{so}(n)$ such that $\mathcal{L}(u_i \wedge u_j) = (\lambda_i + \lambda_j) u_i \wedge u_j$.

Proposition 12. Given $L \in M(n, \mathbb{R})$ define the linear map $\mathcal{L} : M(n, \mathbb{R}) \to M(n, \mathbb{R})$ by $\mathcal{L}(Z) = ZL + LZ$. If L is symmetric and non-negative definite of rank at least $n-1$, then $\mathcal{L}$ is an isomorphism and $\langle Z_1, Z_2 \rangle := \frac{1}{2} Tr(\mathcal{L}(Z_1)Z_2^\dagger)$ is a left-invariant Riemannian metric on $SO(n)$. The tensor B for this metric is

$$B(Z_1, Z_2) = \left[LZ_1, Z_2\right] L^{-1}$$

for all $Z_1, Z_2 \in \mathfrak{so}(n)$.

Proof. Let $\lambda_1, \ldots, \lambda_l$ be the distinct eigenvalues and $V_1, \ldots, V_l$ the respective eigenspaces of L. Let $\pi_j : \mathbb{R}^n \to V_j$ denote the orthogonal projections. Then $\pi_j L = L\pi_j = \lambda_j \pi_j$. It suffices to show that $\mathcal{L}$ has trivial kernel. Thus suppose $\mathcal{L}(Z) = 0$. Then for all i, j,

$$0 = \pi_i \mathcal{L}(Z) \pi_j = \pi_i L Z \pi_j + \pi_i Z L \pi_j = (\lambda_i + \lambda_j) \pi_i Z \pi_j.$$

But $\lambda_i + \lambda_j > 0$ by the assumptions on L so all blocks $\pi_i Z \pi_j$ are zero, hence $Z = 0$. The expression for B follows from $\mathrm{Tr}\left([Z_2, Z_3] L Z_1^\dagger\right) = \mathrm{Tr}\left([LZ_1, Z_2] L^{-1} Z_3^\dagger\right)$. □

3.3.2 Kinetic energy metrics on $SE(n)$ for rigid bodies in $\mathbb{R}^n$

The left-invariant metrics on $G = SE(n)$ of interest here are derived from mass distributions on the rigid body. Let $B \subset \mathbb{R}^n$ denote the body in its *reference configuration*. The position of *material point* $b \in B$ in the configuration $g = (A, a) \in G = SO(n) \times \mathbb{R}^n$ is $\Phi(g, b) := Ab + a$. We

call $\Phi : G \times B \to \mathbb{R}^n$ the *position map* and use the alternative notations $\Phi(g, b) = g(b) = \Phi_b(g)$ as convenience dictates. For now (until we consider collisions shortly) B may be any measurable set with a finite (positive) measure μ defining its mass distribution. Recall from Section 3.1.3 that $m := \mu(B)$ is the mass of the body and the first moment of μ is 0. When considering the motion of several bodies, we assume that the center of mass of each of them in the standard configuration is at 0.

Elements of $\mathfrak{g} = \mathfrak{so}(n) \times \mathbb{R}^n$ will be written in the form $\xi = (Z, z)$. Let L_g and R_g denote left and right-multiplication by g. We will very often use the identification $G \times \mathfrak{g} \cong TG$ given by $(g, \xi) \mapsto (dL_g)_e \xi$. Each $v \in T_g G$ gives rise to the map $V_v : B \to \mathbb{R}^n$ defined by $V_v(b) = (d\Phi_b)_q \, v$, which is the velocity of b in state (g, v). The *kinetic energy Riemannian metric* on G is defined so that the inner product of $u, v \in T_g G$ is given by

$$\langle u, v \rangle_g = \int_B V_u(b) \cdot V_v(b) \, d\mu(b).$$

Proposition 13. The Riemannian metric on $SE(n)$ associated to the mass distribution μ is invariant under left-translations.

Proof. To see this, note first that

$$V_v(b) = \left. \frac{d}{ds} \right|_{s=0} g e^{s\xi} b = \left. \frac{d}{ds} \right|_{s=0} \left(A e^{sZ} b + A \int_0^s e^{tZ} z \, dt + a \right) = A \left(Zb + z \right).$$

Here we have used the form of the exponentiation in $SE(n)$ given in Section 3.2.1. Therefore, as A leaves invariant the standard inner product in $\mathbb{R}^n$,

$$V_u(b) \cdot V_v(b) = \left(Z^u b + z^u \right) \cdot \left(Z^v b + z^v \right)$$

and so $\langle (dL_g)_e \xi, (dL_g)_e \eta \rangle_g = \langle \xi, \eta \rangle_e$. □

Recall the *inertia matrix* L introduced in Section 3.1.3.

Proposition 14. The matrix L associated to mass distribution μ satisfies:

1. For arbitrary $n \times n$ matrices Z_1 and Z_2, $\int_B (Z_1 b) \cdot (Z_2 b)\, d\mu(b) = Tr\left(Z_1 L Z_2^\dagger\right)$.

2. If $A \in SO(n)$, the inertia matrix of the rotated body gB is $L^A := ALA^\dagger$.

3. $L = \lambda I$ if μ is $SO(n)$-invariant. If μ is uniform on a ball of radius R, $\lambda = (n+2)^{-1} R^2$.

Proof. These are obtained by elementary calculations. □

Let $\mathcal{L}(Z) = LZ + ZL$ where, from now on, L is an inertia matrix.

Corollary 3. The kinetic energy Riemannian metric can be written in the form

$$\langle u, v \rangle_g = m\left[\frac{1}{2} Tr\left(\mathcal{L}(Z_u) Z_v^\dagger\right) + z_u \cdot z_v\right] \tag{3.3.1}$$

where $u, v \in T_g G$ and their left-translates to $\mathfrak{g}$ are indicated by (Z_u, z_u) and (Z_v, z_v).

Proposition 15 (Tensor B for $\mathfrak{se}(n)$)**.** Let $SE(n)$ be given the left-invariant Riemannian metric associated to the inertia matrix L. Then

$$B((Z_1, z_1), (Z_2, z_2)) = \left(\left([LZ_1, Z_2] - \frac{1}{2} z_1 \wedge z_2\right) L^{-1}, -Z_2 z_1\right).$$

Proof. Observe that

$$\begin{aligned}
\langle [(Z_2, z_2), (Z_3, z_3)], (Z_1, z_1) \rangle &= \langle ([Z_2, Z_3], Z_2 z_3 - Z_3 z_2), (Z_1, z_1) \rangle \\
&= m\left\{\mathrm{Tr}\left([Z_2, Z_3] L Z_1^\dagger\right) + (Z_2 z_3 - Z_3 z_2) \cdot z_1\right\} \\
&= m\mathrm{Tr}\left(\left([LZ_1, Z_2] - \frac{1}{2} z_1 \wedge z_2\right) L^{-1} L Z_3^\dagger\right) - m(Z_2 z_1) \cdot z_3 \\
&= \left\langle \left(\left([LZ_1, Z_2] - \frac{1}{2} z_1 \wedge z_2\right) L^{-1}, -Z_2 z_1\right), (Z_3, z_3) \right\rangle.
\end{aligned}$$

The claimed identity now follows from the definition on B. □

We note that if $L = \lambda I$, then $B((Z,z),(Z,z)) = (0, -Zz)$.

Proposition 16. Give $G = SE(n)$ the left-invariant Riemannian metric defined by a mass distribution on the rigid body B with inertia matrix L and mass m. Let $\xi = (W,w) \in \mathfrak{g}$ and $(g,v) \in TG$ where $g = (A,a)$ and $v = (dL_g)_e(Z,z)$. Then

$$\mathcal{P}^{\mathfrak{g}}(g,v)(\xi) = \frac{1}{2} m\, Tr\left\{ \left(Ad_A \mathcal{L}(Z) + x_c \wedge v_c \right) W^{\dagger} \right\} + m v_c \cdot w.$$

Here $x_c = a$ is the position of the center of mass of the body in configuration g and $v_c \coloneqq Az$ is the velocity of the center of mass for the given state (g,v).

Proof. This is a straightforward computation based on the definition of $\mathcal{P}^{\mathfrak{g}}$ and the expression of the Riemannian metric given in Corollary 3. □

Let $\langle \cdot, \cdot \rangle_{\mathfrak{g}}$ be the left-invariant inner product on $\mathfrak{g}$ given by

$$\langle (Z,z),(W,w) \rangle_{\mathfrak{g}} \coloneqq \mathrm{Tr}(ZW^{\dagger}) + z \cdot w. \tag{3.3.2}$$

Then, with the notation of Proposition 16,

$$\mathcal{P}^{\mathfrak{g}}(g,v)(\xi) = m \left\langle \left(\frac{1}{2} \left(\mathrm{Ad}_A \mathcal{L}(Z) + x_c \wedge v_c \right), v_c \right), (W,w) \right\rangle_{\mathfrak{g}} \tag{3.3.3}$$

3.3.3 Singular force fields and impulses

Let M be the configuration manifold of a mechanical system with the kinetic energy Riemannian metric and material body B with mass distribution measure μ.

A *force field* on M is a bundle map $F : TM \to T^*M$ possibly depending on time, although we omit explicit reference to the time variable. So if $v \in T_qM$, then $F(q,v) \in T_q^*M$. Forces

acting on B typically arise from a $\mathbb{R}^n$-valued (possibly time dependent) measure φ on B, parametrized by TM, called the *force distribution*. From such a measure we define the force field $F(q,v) \in T_q^*M$ such that for each $u \in T_qM$,

$$F(q,v)(u) = \int_B V_u(b) \cdot d\varphi_{q,v}(b).$$

We are interested in cases where $\varphi_{q,v}$ is singular, supported on a single point of B.

Definition 8 (Newton's equation)**.** Newton's equation of motion of the (unconstrained) mechanical system with configuration manifold $(M, \langle\cdot,\cdot\rangle)$ and force field F is

$$\frac{\nabla}{dt}\mathcal{P}(q,\dot{q}) = F(t,q,\dot{q}).$$

Proposition 17. Give $M = SE(n)$ a left-invariant Riemannian metric and let F be a force field on M. Let $F^\#$ be the dual field, so $F(t,q,v)(u) = \langle F^\#(t,q,v), u\rangle_q$. Then

$$\frac{\nabla}{dt}\mathcal{P}(g,\dot{g}) = F \Leftrightarrow \frac{\nabla \dot{g}}{dt} = F^\# \Leftrightarrow \dot{w} - B(w,w) = (dL_{g^{-1}})_g F^\# \Leftrightarrow \frac{d}{dt}\mathcal{P}^{\mathfrak{g}}(g,\dot{g}) = R_g^* F.$$

Proof. These are consequences of Propositions 8 and 11. □

When $M = SE(n)$, it is useful to regard the force field as a Lie algebra-valued by left-translating each force vector to T_eG. We define

$$\mathcal{F}(t,g,\dot{g}) \coloneqq (Y(t,g,\dot{g}), y(t,g,\dot{g})) \coloneqq (dL_g)_e^{-1} F^\#(t,g,\dot{g}) \in \mathfrak{g}.$$

Then, using the notation of 3.3.2,

$$\begin{aligned} \langle \mathcal{F}, \mathrm{Ad}_{g^{-1}}\xi\rangle_{\mathfrak{g}} &= \frac{1}{2}\mathrm{Tr}\left[\left(\mathrm{Ad}_A \mathcal{L}(Y) + x_c \wedge Ay\right) W^\dagger\right] + (Ay)\cdot w \\ &= \left\langle \left(\frac{1}{2}\left(\mathrm{Ad}_A \mathcal{L}(Y) + x_c \wedge Ay\right), Ay\right), (W,w)\right\rangle_{\mathfrak{g}} \end{aligned} \tag{3.3.4}$$

Proposition 3 follows from these remarks, keeping in mind that $v_c = Az$ and $x_c = a$.

Let the force be applied on a single point $Q = Q(t, g, \dot{g}) \in \mathbb{R}^n$, so that φ is supported on Q. Let $u = (dR_g)\xi \in T_qM$, $\xi = (W, w) \in \mathfrak{g}$. Note that we are using right-translation here since we wish to evaluate the last of the equivalent equations of Proposition 17 on the Lie algebra element ξ. Then there exists $\mathcal{I} = \mathcal{I}(t, g, \dot{g})$ depending only of F such that

$$F(t, g, \dot{g})(u) = \int_B V_u(b) \cdot d\varphi_{g,\dot{g}}(b) = V_u(g^{-1}Q) \cdot \mathcal{I} = (WQ + w) \cdot \mathcal{I} = \frac{1}{2}\mathrm{Tr}((\mathcal{I} \wedge Q)W^\dagger) + \mathcal{I} \cdot w.$$

This gives

$$F(t, g, \dot{g})(u) = \left\langle \left(\frac{1}{2} Q \wedge \mathcal{I}, \mathcal{I} \right), (W, w) \right\rangle_{\mathfrak{g}}.$$

It follows from the expression 3.3.4 of F that

$$\mathrm{Ad}_A \mathcal{L}(Y) + x_c \wedge Ay = Q \wedge \mathcal{I}$$
$$Ay = \mathcal{I}.$$

Writing $f_c := Ay$, this is equivalent to $f_c = \mathcal{I}$ and $\mathrm{Ad}_A \mathcal{L}(Y) = (Q - x_c) \wedge f_c$. Therefore, the equation of motion becomes

$$m\dot{v}_c = f_c$$
$$m\,\mathrm{Ad}_A \left(\mathcal{L}\dot{Z} - [\mathcal{L}Z, Z] \right) = (Q - x_c) \wedge f_c$$

proving Proposition 4.

Our informal discussion of the idea of impulse from earlier and the above remarks now give the expression $((Q - x_c) \wedge \mathcal{I}_c, \mathcal{I}_c) \in \mathfrak{g}$ for the change in momentum due to singular forces applied to Q. This gives the following.

Proposition 18 (Change in momentum due to impulsive forces)**.** If the rigid body with

mass m, inertia matrix L and associated Lie algebra map $\mathcal{L}$, is subject to an impulsive force concentrated at point $Q \in \mathbb{R}^n$ at a given time, then momentum changes discontinuously according to

$$\begin{aligned} mv_c^+ - mv_c^- &= \mathcal{I}_c \\ m\, Ad_A \mathcal{L}(Z^+ - Z^-) &= (Q - x_c) \wedge \mathcal{I}_c \end{aligned} \tag{3.3.5}$$

for some vector $\mathcal{I}_c \in \mathbb{R}^n$ depending on the state of the body. As before, $x_c = a$ indicates the center of mass of the body in configuration $g = (A, a)$, $v^\pm = Az^\pm$ are the velocities of the center of mass immediately prior to and after the application of the impulse, and $(g, (Z^\pm, z^\pm))$ are the pre- and post-impulse states of the body.

3.3.4 Several bodies and momentum conservation

If the mechanical system consists of several unconstrained rigid bodies, $B_1, \ldots, B_k$ (in reference configuration) subject to forces $F_j(q, v)$, $j = 1, \ldots k$, the configuration manifold M is a subset of the product $G \times \cdots \times G$, with one copy of $G = SE(n)$ for each body. We consider for now only motion in the interior of M.

We say that forces are *internal* to the system if they are somehow due to the influence of the bodies on each other. More specifically, we use term 'internal' when $F_j = \sum_{i \neq j} F_{ij}$ and the $F_{ij} = F_{ij}(q, v)$—the force body i exerts on body j in state (q, v)—satisfies the property of *action-reaction*: $F_{ij} = -F_{ji}$. If the forces are derived from, possibly singular, measures $\varphi_{q,v}(j, b|i, b')$ on $B_i \times B_j$ so that

$$F_{ij}(q, v)(u) = \int_{B_i} \int_{B_j} V_u(b') \cdot d\varphi_{q,v}(j, b'|i, b),$$

then the action-reaction property, expressed in terms of φ, means that

$$d\varphi(i, b|j, b') = -d\varphi(j, b'|i, b)$$

for almost every b, b' (with respect to φ). Newton's equation applied to body j is then

$$\frac{\nabla}{dt}\mathcal{P}_j(q,\dot{q}) = \sum_{i\neq j} F_{ij}(t,q,\dot{q})$$

and the total momentum $\mathcal{P}(q,\dot{q}) = \sum_j \mathcal{P}_j(q,\dot{q})$ is conserved:

$$\frac{\nabla}{dt}\mathcal{P}(q,\dot{q}) = \sum_j \sum_{i\neq j} F_{ij}(t,q,\dot{q}) = 0.$$

Another way to interpret the notion of forces internal to the system is to assume that the total work the F_i do along a rigid motion of the entire system, that is, the work along a path in M of the form $\gamma(t) := e^{t\xi}q = (e^{t\xi}g_1, \ldots, e^{t\xi}g_k)$ is zero. The total work is then

$$\begin{aligned}
0 &= \int_a^b \sum_j F_j(\gamma(t),\gamma'(t))(\gamma_j'(t))\,dt \\
&= \sum_j \int_a^b (R_g^* F_j(\gamma(t),\gamma'(t)))(\xi)\,dt \\
&= \sum_j \int_a^b \frac{d}{dt}\left[\mathcal{P}_j^{\mathfrak{g}}(\gamma_j(t),\gamma_j'(t))(\xi)\right]dt \\
&= \sum_j \mathcal{P}_j^{\mathfrak{g}}(\gamma(b),\gamma'(b))(\xi) - \sum_j \mathcal{P}_j^{\mathfrak{g}}(\gamma(a),\gamma'(a))(\xi)
\end{aligned}$$

and, again, the total momentum (now in the sense of the momentum map on $\mathfrak{g}$) is constant. In this sense, conservation of momentum follows, as expected, from a symmetry property.

Of particular interest here are two bodies that interact through impulses applied to a common point of collision Q. Then for each body, indicated by the index $i = 1, 2$,

$$\begin{aligned}
m_i v_{c,i}^+ - m_i v_{c,i}^- &= \mathcal{I}_{c,i} \\
m_i \operatorname{Ad}_{A_i}\mathcal{L}_i(Z_i^+ - Z_i^-) &= (Q - x_{c,i}) \wedge \mathcal{I}_{c,j}
\end{aligned} \tag{3.3.6}$$

where the impulse vectors satisfy $\mathcal{I}_{c,1} + \mathcal{I}_{c,2} = 0$ by conservation of momentum. Proposition 5

is now a consequence of this observation and of Proposition 18.

3.4 Kinematics of two rigid bodies

The configuration manifold of a system of several (unconstrained) rigid bodies in $\mathbb{R}^n$ is a submanifold with boundary of the product of copies of the Euclidean group $SE(n)$, one copy for each body. The Riemannian metric is then the product of the Riemannian metrics for each single body. Here we focus on the boundary of the configuration manifold of two bodies and certain structures therein.

Let B_1 and B_2 be submanifolds of $\mathbb{R}^n$ of dimension n having smooth boundary and equipped with mass distribution measures μ_1 and μ_2 with masses $m_j := \mu_j(B_j) < \infty$ and zero first moment. The bodies need not be bounded. The configuration manifold M is by definition the closure of

$$M_0 := \{q = (g_1, g_2) \in G \times G : g_1(B_1) \cap g_2(B_2) = \varnothing\} \tag{3.4.1}$$

where $G = SE(n)$. We further assume that each collision configuration $q = (g_1, g_2) \in \partial M$ is such that $g_1(B_1) \cap g_2(B_2)$ consists of a single point.

The definition of M as the closure of M_0 is not a very useful description of M near its boundary. In particular, it is not so clear how to translate geometric information about the boundaries of the B_j into information about the boundary of M. For this purpose we introduce the *extended configuration manifold* M_e defined below.

Let N_j be the boundary of B_j and let ν_j be the outward-pointing unit normal vector field on N_j. By a (positive) *adapted* orthonormal frame at $b \in N_j$ of sign $\epsilon \in \{+, -\}$ we mean a positive orthogonal map $\sigma : \mathbb{R}^n \to T_b\mathbb{R}^n \cong \mathbb{R}^n$ such that $\sigma e_n = \epsilon\nu_j(b)$. Here e_n is the last vector of the standard basis $(e_1, \ldots, e_n)$ of $\mathbb{R}^n$. Hence σ is an element of $SO(n)$ mapping $\mathbb{R}^n$ isometrically to T_bN_j. If σ is an adapted frame and $h \in H := SO(n-1)$, then σh is also an

adapted frame with the same base point as σ. In this way, H acts freely and transitively by right multiplication on the set of adapted frames at any given point of N_j.

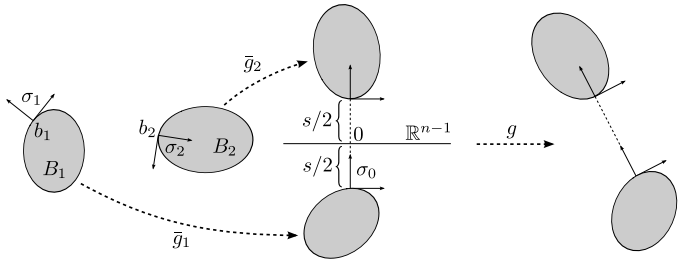

Figure 3.3: Interpretation of the map Ψ. The transformation $\bar{g}_j$ sends body B_j from its standard configuration to the configuration that takes the adapted frame σ_j to the standard frame in $\mathbb{R}^n$, and the point b_j into the line through the origin along e_n a distance $s/2$ from the origin.

We denote by $\mathcal{F}_j^\epsilon$ the principal H-bundle of adapted (positive) orthonormal frames over N_j of sign ϵ. Elements of $\mathcal{F}_j^\epsilon$ will be written (b, σ), where σ is in the fiber $\mathcal{F}_j^\epsilon(b)$. The *extended configuration manifold* is the product $M_e \coloneqq \mathcal{F}_1^+ \times \mathcal{F}_2^- \times G \times [0, \infty)$. We can now define the map $\Psi : M_e \to G \times G$ by $\Psi(b_1, \sigma_1, b_2, \sigma_2, g, s) = (g_1, g_2)$ where

$$g_1 = g\bar{g}_1 = g\left(\sigma_1^{-1}, -\sigma_1^{-1}b_1 - s/2\right) = \left(A\sigma_1^{-1}, a - \frac{s}{2}Ae_n - A\sigma_1^{-1}b_1\right)$$
$$g_2 = g\bar{g}_2 = g\left(\sigma_2^{-1}, -\sigma_2^{-1}b_2 + s/2\right) = \left(A\sigma_2^{-1}, a + \frac{s}{2}Ae_n - A\sigma_2^{-1}b_2\right).$$

The geometric interpretation of Ψ is shown in Figure 3.3. Note that points on the boundary of M correspond under Ψ to points in M_e with coordinate $s = 0$. The groups G and H naturally act on M_e on left and right, respectively:

$$g(b_1, \sigma_1, b_2, \sigma_2, g', s)h \coloneqq (b_1, \sigma_1 h, b_2, \sigma_2 h, gg'h, s).$$

The quotient M_e/H is easily seen to be a smooth manifold and the projection $M_e \to M_e/H$ is

a principal H-bundle. It is also immediate from the definitions that

$$\Psi(gqh) = g\Psi(q)$$

for all $\xi \in M_e$, where the action of G on $G \times G$ is defined by $g(g_1, g_2) = (gg_1, gg_2)$. Therefore, Ψ induces a G-equivariant map

$$\overline{\Psi} : M_e/H \to G \times G.$$

Equivariance means $\overline{\Psi}(gq) = g\overline{\Psi}(q)$. The G-action on M_e admits a smooth cross-section:

$$S := \mathcal{F}_1^+ \times \mathcal{F}_2^- \times \{e\} \times [0, \infty).$$

The G-action on M_e and on M_e/H leaves invariant the coordinate s; in particular, it leaves the boundary of these two manifolds invariant.

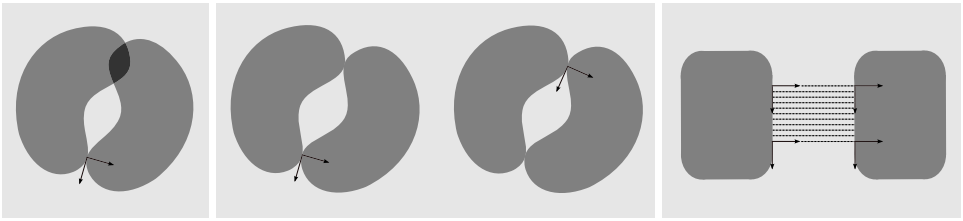

Figure 3.4: For M_e/H to be a good parametrization of M near the boundary some pathologies must be avoided. Far left: a boundary configuration in M_e/H that is not in ∂M; middle pair: two distinct elements of M_e/H corresponding to the same element in ∂M; far right: a curve in M_e that is mapped under $\overline{\Psi}$ to a single point in M.

It is natural to expect that under reasonable assumptions the restriction of $\overline{\Psi}$ to a neighborhood of the boundary of M_e/H will be a diffeomorphism onto a neighborhood of the boundary of M, thus providing a useful parametrization for the purpose of understanding collisions. Figure 3.4 shows some of the situations that must be avoided. We give shortly a few sets of sufficient conditions for $\overline{\Psi}$ to be a local diffeomorphism, but our immediate goal is

to explore M_e, M_e/H, and their boundaries a little further.

3.4.1 The tangent bundle of ∂M_e and $\partial(M_e/H)$

Recall that the shape operator of a hypersurface $N \subset \mathbb{R}^n$ with unit normal vector field ν at a point $b \in N$ is the linear map $S_b : T_bN \to T_bN$ defined by $v \mapsto -D_v\nu$ where D_v is the Levi-Civita connection for the standard Euclidean metric in $\mathbb{R}^n$. We write $\nabla_v X := \Pi_b D_v X$ for a tangent vector field X on N, where Π_b is the orthogonal projection from $\mathbb{R}^n$ to T_bN. This is the Levi-Civita connection on N for the induced metric. Let (b, σ) be a point in the adapted frame bundle $\mathcal{F}^\epsilon$ over N and (v, ζ) a tangent vector to $\mathcal{F}^\epsilon$ at (b, σ). Then, by differentiating $\nu(\gamma(t)) = -(-1)^\epsilon \sigma(t) e_n$, where $(\gamma(t), \sigma(t))$ is a smooth curve representing (v, ζ) at $(b, \sigma) = (\gamma(0), \sigma(0))$, we obtain

$$S_b(v) = (-1)^\epsilon \zeta e_n = -\zeta \sigma^{-1} \nu(b) = -\sigma V \sigma^{-1} \nu(b)$$

where $V := \sigma^{-1}\zeta$ can be regarded as an element of $\mathfrak{so}(n)$ just as σ is viewed as an element of $SO(n)$. The tangent bundle of $\mathcal{F}^\epsilon$ for any smooth hypersurface N has now the following description. Let $(b, \sigma) \in \mathcal{F}^\epsilon$. Then

$$T_{(b,\sigma)}\mathcal{F}^\epsilon \cong \left\{ (v, V) \in T_bN \times \mathfrak{so}(n) : S_b(v) = -\sigma V \sigma^{-1} \nu(b) \right\}.$$

As before, we use the canonical identification $TG \cong G \times \mathfrak{g}$ and write elements of $\mathfrak{g}$ in the form $(Z, z) \in \mathfrak{so}(n) \times \mathbb{R}^n$. The shape operator of N_j will be written $S^{(j)}$. We omit the superscript when it is clear from the context to which body the operator is associated. Then the tangent

space of M_e at a point $q = (b_1, \sigma_1, b_2, \sigma_2, g, s)$ is given by

$$T_q M_e = \Big\{(v_1, V_1, v_2, V_2, Z, z, \varrho) : v_j \in T_{b_j} N_j, V_j \in \mathfrak{so}(n), (Z, z) \in \mathfrak{g}, \varrho \in \mathbb{R}, \\ S_{b_j} v_j = -\sigma_j V_j \sigma_j^{-1} \nu_j(b_j), j = 1, 2\Big\}.$$

Tangent spaces to ∂M_e consist of those vectors for which $\varrho = 0$.

Let G_q and H_q represent the orbits through $q \in \partial M_e$ of the (right and left, respectively) actions of G and H on M_e. The tangent spaces at q of the respective orbits will be written $\mathfrak{g}_q$ and $\mathfrak{h}_q$. Then

$$\mathfrak{g}_q = \{(0, 0, 0, 0, Z, z, 0) : (Z, z) \in \mathfrak{g}\} \text{ and } \mathfrak{h}_q = \{(0, Y, 0, Y, Y, 0, 0) : Y \in \mathfrak{h}\}. \tag{3.4.2}$$

At any $q \in \partial M_e$ the differential of Ψ is

$$d\Psi_q(v_1, V_1, v_2, V_2, Z, z, \varrho) = (Z_1, z_1, Z_2, z_2),$$

where, denoting $\mathrm{Ad}_\sigma(W) := \sigma W \sigma^{-1}$,

$$\begin{aligned} Z_j &= \mathrm{Ad}_{\sigma_j}(Z - V_j) \\ z_j &= \sigma_j z - \mathrm{Ad}_{\sigma_j}(Z - V_j) b_j - v_j - \varrho \nu_j(b_j). \end{aligned} \tag{3.4.3}$$

The next proposition contains as a special case Proposition 2. It uses the notation

$$S_j := \sigma_j^{-1} S_{b_j}^{(j)} \sigma_j : \mathbb{R}^{n-1} \to \mathbb{R}^{n-1}$$

for any $q = (b_1, \sigma_1, b_2, \sigma_2, g, s)$. We allow s to be non-zero, in which case $S^{(j)}$ is the shape operator of the level hypersurface of M_e/H corresponding to value s.

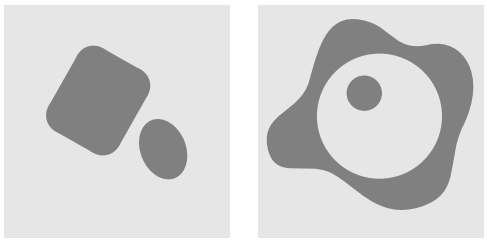

Figure 3.5: Situations for which Proposition 19 applies.

Proposition 19. The map $\Psi : M_e \to G \times G$ is a submersion from a neighborhood $\mathcal{U}$ of the boundary of M_e onto a neighborhood of the boundary of M if any of the following conditions involving $\mathcal{U}$ and the shape operators holds.

1. $S_1 + S_2$ is non-singular at all points of ∂M_e. In this case, $\mathcal{U}$ is a neighborhood of ∂M_e where this non-singular condition holds.

2. $\mathcal{U}$ is a neighborhood of ∂M_e where one of the S_j is non-singular and $S_1 + S_2 - sS_1S_2$ is also non-singular.

3. If the two bodies are convex and the boundary of one of them has non-vanishing Gauss-Kronecker curvature so that S_j is everywhere non-singular on N_j for some j, then $\mathcal{U} = M_e$.

In each case $\mathcal{U}$ is G-invariant, the kernel of $d\Psi_q$ is $\mathfrak{h}_q$ at each $a \in \mathcal{U}$, and $\Psi|_{\mathcal{U}} : \mathcal{U} \to \Psi(\mathcal{U})$ is a principal H-bundle. In addition, the boundary of M is a smooth submanifold and there are smooth functions $b_j : \partial M \to N_j$, $j = 1, 2$, such that $b_1(q), b_2(q)$ are the unique points on the respective bodies that are brought into contact in collision configuration q.

Proof. By counting dimensions we see that $\mathcal{U}$ should be a neighborhood of the boundary of M_e where the kernel of $d\Psi_q$ is $\mathfrak{h}_q$. It follows from equations 3.4.2 and 3.4.3 above that this kernel contains $\mathfrak{h}_q$. We show equality under the conditions of item (2), the other cases being similar. Say that S_2 is non-singular. From the explicit form of $d\Psi_q$ given in 3.4.3 we see that

$\xi = (v_1, V_1, v_2, V_2, Z, z, \varrho)$ lies in that kernel if and only if $Z = V_1 = V_2$ and $z = \sigma_j^{-1} v_j - (-1)^j \frac{s}{2} Z e_n$ for $j = 1, 2$. Observe that $\sigma_j^{-1} v_j$ and $Z e_n$ lie in $\mathbb{R}^{n-1}$, which is orthogonal to e_n. Hence $\varrho = 0$. Keeping in mind $\sigma_j V_j e_n = (-1)^j S_{b_j} v_j$, we obtain

$$-S_1 z = Z e_n - \frac{s}{2} S_1 Z e_n$$
$$S_2 z = Z e_n - \frac{s}{2} S_2 Z e_n.$$

From this we conclude that $[S_1 + S_2 - s S_1 S_2] S_2^{-1} Z e_n = 0$ which, under the conditions of (2) implies that $Z e_n = 0$. Since S_2 is non-singular, this also implies that $z = 0$ and $v_j = 0$. Therefore, $\xi = (0, Z, 0, Z, Z, 0, 0)$, where $Z \in \mathfrak{h}$ since $Z e_n = 0$. That $\Psi|_{\mathcal{U}}$ is a principal H-bundle is now easy. G-equivariance of Ψ implies that $\mathcal{U}$ is G-invariant. □

Assuming for simplicity that $\mathcal{U}$ of Proposition 19 is all of M_e (we only need what follows on some neighborhood of the boundary of M_e), it is useful to know whether the principal bundle $M_e \to M$ admits a G-invariant connection since the associated horizontal subspace $\mathcal{H}_q$ can then serve as a proxy for the tangent space of $T_{\bar{q}} M$, without having to go to the quotient. A principal H-connection on M_e is given by a one-form ω taking values in $\mathfrak{h}$ and satisfying the properties:

1. $\omega_q(Y_q) = Y \in \mathfrak{h}$, where Y_q is the vector induced by the infinitesimal action of $\mathfrak{h}$;
2. $h^* \omega = \mathrm{Ad}_{h^{-1}} \circ \omega$.

Let Π be the orthogonal projection from $\mathbb{R}^n$ to $\mathbb{R}^{n-1} = e_n^{\perp}$.

Proposition 20. For any real constants c_1, c_2, c_3 such that $c_1 + c_2 + c_3 = 1$ the $\mathfrak{h}$-valued one-form ω on M_e given by

$$\omega_q(v_1, V_1, v_2, V_2, Z, z, \varrho) = c_1 \Pi V_1 \Pi + c_2 \Pi V_2 \Pi + c_3 \Pi Z \Pi$$

is a G-invariant H-connection on M_e.

Proof. This is a simple check. □

Let us choose $\omega_q(\xi) := \Pi V_1 \Pi$ and denote by $\mathcal{H}_q$ the horizontal subspace defined by this choice of connection form. Recall the maps S_j on $\mathbb{R}^{n-1}$ for $q = (b_1, \sigma_1, b_2, \sigma_2, g, 0) \in \partial M_e$ given by $S_j = \sigma_j^{-1} S_{b_j} \sigma_j$. The b_j are determined uniquely from $\Psi(q)$ and the σ_j are determined uniquely up to an overall common element of H acting on the right.

Proposition 21. Let $q = (b_1, \sigma_1, b_2, \sigma_2, g, 0) \in \partial M_e$ and suppose that $S_1 + S_2$ is invertible. Then $d\Psi_q$ maps $\mathcal{H}_q$ isomorphically onto $\mathfrak{g} \times \mathfrak{g}$ and

$$d\Psi_q\left(\mathcal{H}_q \cap T_q(\partial M_e)\right) = \{(Z_1, z_1, Z_2, z_2) \in \mathfrak{g} \times \mathfrak{g} : \nu_1 \cdot (Z_1 b_1 + z_1) + \nu_2 \cdot (Z_2 b_2 + z_2) = 0\}$$

where $\nu_1 = \nu_1(b_1)$ and $\nu_2 = \nu_2(b_2)$ are the unit normal vectors.

Proof. The proof is elementary, but we show the main point. Let $\bar{\xi} = (Z_1, z_1, Z_2, z_2) \in \mathfrak{g} \times \mathfrak{g}$. We wish to show the existence of a unique $\xi = (v_1, V_1, v_2, V_2, Z, z, \varrho) \in \mathcal{H}_q$ that is sent to $\bar{\xi}$ under $d_q\Psi$. The components of ξ satisfy:

$$\sigma_j^{-1} v_j \in \mathbb{R}^{n-1},\ \Pi V_j \Pi = 0,\ (Z, z) \in \mathfrak{g},\ \varrho \in \mathbb{R},\ S_j \sigma_j^{-1} v_j = (-1)^j V_j e_n.$$

and Z_j, z_j are related to theses quantities by

$$Z_j = \mathrm{Ad}_{\sigma_j}(Z - V_j), \quad z_j = \sigma_j z - \mathrm{Ad}_{\sigma_j}(Z - V_j) b_j - v_j - \varrho \nu_j(b_j).$$

Writing $g = (A, a)$ and $\Psi(q) = (g_1, g_2)$, we have $g_j = (A\sigma_j^{-1}, a - A\sigma_j^{-1} b_j)$. Note that

$$\sigma_1^{-1} v_1 - \sigma_2^{-1} v_2 + 2\varrho e_n = \sigma_2^{-1}(Z_2 b_2 + z_2) - \sigma_1^{-1}(Z_1 b_1 + z_1) \tag{3.4.4}$$

from which we obtain ϱ and $\sigma_1^{-1}v_1 - \sigma_2^{-1}v_2 \in \mathbb{R}^{n-1}$ in terms of the Z_j and z_j. Also

$$\begin{aligned}\left(\mathrm{Ad}_{\sigma_2^{-1}}Z_2 - \mathrm{Ad}_{\sigma_1^{-1}}Z_1\right)e_n &= (V_2 - V_2)e_n \\ &= S_1\sigma_1^{-1}v_1 + S_2\sigma_2^{-1}v_2 \\ &= (S_1+S_2)\sigma_1^{-1}v_1 + S_2(\sigma_2^{-1}v_2 - \sigma_1^{-1}v_1) \\ &= (S_1+S_2)\sigma_1^{-1}v_1 + S_2\Pi\left\{\sigma_2^{-1}(Z_2b_2+z_2) - \sigma_1^{-1}(Z_1b_1+z_1)\right\}\end{aligned}$$

from which we obtain $\sigma_1^{-1}v_1$ in terms of the Z_j and z_j under the assumption that $S_1 + S_2$ is invertible. From Proposition 7, item (8), we deduce

$$V_1 = \Pi V_1 \Pi + e_n \wedge V_1 e_n = e_n \wedge V_1 e_n = -e_n \wedge S_1\sigma_1^{-1}v_1$$

so that V_1 is also uniquely determined by the Z_j and z_j. From

$$V_2 - V_1 = \mathrm{Ad}_{\sigma_2^{-1}}Z_2 - \mathrm{Ad}_{\sigma_1^{-1}}Z_1$$

we obtain V_2 uniquely and from the above 3.4.4 we obtain v_2 uniquely. From these we easily obtain Z and z as well, proving the first part of the proposition. The second part follows from the observation that a vector is tangent to ∂M if and only if $\varrho = 0$. □

3.4.2 The non-slipping, rolling, and diagonal subbundles

Let $\gamma(t)$ be a smooth curve in ∂M_e such that $q = \gamma(0) = (b_1, \sigma_1, b_2, \sigma_2, g, 0)$. We omit the variable s, which is set to 0 for a boundary point. Let $(\gamma_1(t), \gamma_2(t))$ be the image of γ under Ψ and write $\gamma_j(0) = g_j$, $\bar{q} := (g_1, g_2)$, where $g_j = (A_j, a_j)$ and $g = (A, a)$. Denote the components of the infinitesimal motion in M_e by

$$\xi := \gamma'(0) = (v_1, V_1, v_2, V_2, Z, z),$$

omitting $\varrho = 0$. The two bodies in configuration q are in contact at $g_1(b_1) = g_2(b_2)$.

Definition 9 (Non-slipping and non-twisting conditions)**.** The infinitesimal motion $\xi \in T_q M_e$ is said to satisfy the *non-slipping condition* if the velocities of the material points b_j at the contact configuration are equal. It is said to satisfy the *non-twisting condition* if the tangent planes to N_j at b_j do not rotate relative to each other under ξ.

We now derive an explicit expression for these conditions. The infinitesimal motion of B_j is given by $\xi_j := (Z_j, z_j) \in \mathfrak{g}$, which is obtained from $d\Psi_q \xi$. We know that

$$\begin{aligned} Z_j &= \mathrm{Ad}_{\sigma_j}(Z - V_j) \\ z_j &= \sigma_j z - \mathrm{Ad}_{\sigma_j}(Z - V_j) b_j - v_j. \end{aligned}$$

Due to Proposition 13 $V_{\xi_j}(b_j) = A_j(Z_j b_j + z_j) = A(z - \sigma_j^{-1} v_j)$. The non-slipping condition, $V_{\xi_1}(b_1) = V_{\xi_2}(b_2)$, then reduces to

$$\sigma_1^{-1} v_1 = \sigma_2^{-1} v_2. \tag{3.4.5}$$

Turning now to the non-twisting condition, let u_j be a tangent vector to N_j at b_j such that, in the contact configuration given by q, is sent to a common vector, for $j = 1, 2$, in the plane of contact. Thus $A_1 u_1 = A_2 u_2$. The infinitesimal rotation of $A_j u_j$ at the point of contact is

$$A_j Z_j u_j = A(Z - V_j)\sigma_j^{-1} u_j.$$

The orthogonal projection to the plane of contact is $A\Pi A^{-1}$, recalling that Π is the orthogonal projection to $\mathbb{R}^{n-1}$. (It may be helpful to keep in mind Figure 3.3.) Because $A = A_j \sigma_j$, the non-twisting condition takes the form $\Pi V_1 \Pi = \Pi V_2 \Pi$ and since $\Pi V_1 \Pi = 0$ holds for horizontal vectors, $\Pi V_j \Pi = 0$ for $j = 1, 2$.

Now let $\Psi(q) = (g_1, g_2)$, $g_j = (A_j, a_j)$ and $\bar{\xi} = (Z_1, z_1, Z_2, z_2) = d\Psi_q \xi$. The non-slipping

condition expressed in terms of $\bar{\xi}$ becomes

$$A_1[Z_1 b_1 + z_1] = A_2[Z_2 b_2 + z_2]$$

and the non-twisting condition becomes

$$\mathrm{Ad}_{A_j} Z_j = W + \nu_j(b_j) \wedge w_j$$

for a $W \in \mathfrak{so}(n)$ independent of h and $w_j \in T_{b_j} N_j$.

Definition 10 (Non-slipping, rolling, and diagonal subbundle)**.** The *non-slipping subbundle* of $T(\partial M)$ consists of all tangent vectors satisfying the non-slipping condition. The *rolling subbundle* of $T(\partial M)$ consists of all tangent vectors satisfying both the non-slipping and non-twisting conditions. The *diagonal subbundle* of $T(\partial M)$ is the tangent bundle to the orbits of the action of G on ∂M defined by $g(g_1, g_2) = (gg_1, gg_2)$. We denote these three subbundles, respectively, $\mathfrak{S}, \mathfrak{R}, \mathfrak{D}$. We refer to these collectively as the *kinematic subbundles* of $T(\partial M)$. Notice that $\mathfrak{D}_{\Psi(q)} = \mathfrak{g}_q$, using previous notation.

Starting from this definition rather than Definition 2, the content of the latter becomes a statement, which is proved by the above remarks.

3.5 Collision maps

Let now $M \subset G \times G$ be the configuration manifold of two rigid bodies in $\mathbb{R}^n$, where $G = SE(n)$. By condition 2 of Proposition 19 M has smooth boundary and boundary points represent configurations in which the bodies are in contact at a single point. Let the state of the bodies before and after collision be given by the element of $T_q M$, $q \in \partial M$, represented by

$$(Z_1^{\pm}, z_1^{\pm}, Z_2^{\pm}, z_2^{\pm}) \in \mathfrak{g} \times \mathfrak{g}.$$

Here the sign '+' indicates post-collision velocities and '−' pre-collision velocities. We obtained in 3.3.6 a condition on the pre- and post-collision velocities due to impulsive forces that act at a single point of the body. We restate it here. Let the common point of contact be $Q = A_j b_j + a_j$, where b_j is the material point in standard body configuration which corresponds to Q in configuration $g_j = (A_j, a_j)$, where $a_j = x_{cj}$ is the center of mass of body j in the given configuration. Then we obtain from expression 3.3.6:

$$\begin{aligned} z_j^+ &= z_j^- + u_j \\ Z_j^+ &= Z_j^- + \mathcal{L}_j^{-1}(b_j \wedge u_j) \end{aligned} \tag{3.5.1}$$

where $u_j = A_j^{-1}\mathcal{I}_{cj}/m_j$. We should add to these equations $\mathcal{I}_{c1} + \mathcal{I}_{c2} = 0$ for conservation of (linear) momentum.

Proof of Theorem 1. A simple dimension count gives $\dim \mathfrak{C}_q = n$ and $\dim \mathfrak{S}_q = 2\dim \mathfrak{g} - n$ so that the sum of the two dimensions equals $\dim T_q M$. Therefore, it suffices to show that these subspaces are orthogonal. The Riemannian metric on M is the restriction of the product metric on $G \times G$ (each factor having a possibly different metric as the bodies may have different mass distributions.) Explicitly, let $u, v \in T_q M$ and write

$$v = ((Y_1, y_1), (Y_2, y_2)), \;\; w = ((Z_1, z_1), (Z_2, z_2)).$$

Then

$$\langle v, w \rangle_q = \sum_j m_j \left[\frac{1}{2} \mathrm{Tr}\left(\mathcal{L}_j(Y_j) Z_j^\dagger \right) + y_j \cdot z_j \right]. \tag{3.5.2}$$

Now consider the vectors

$$\begin{aligned} v &= ((\mathcal{L}_1^{-1}(b_1 \wedge u_1), u_1), (\mathcal{L}_2^{-1}(b_2 \wedge u_2), u_2)) \in \mathfrak{C}_q \\ w &= ((A_1^{-1} Z_1 A_1, A_1^{-1} z^* - A_1^{-1} Z_1 A_1 b_1), (A_2^{-1} Z_2 A_2, A_2^{-1} z^* - A_2^{-1} Z_2 A_2 b_2)) \in \mathfrak{S}_q \end{aligned}$$

where $Z_j = Z - V_j$ and $z^* = z - z'$. Observe that

$$\mathrm{Tr}\left((b_j \wedge u_j)(A_j^{-1} Z_j A_j)^\dagger)\right) = 2u_j \cdot (A_j^{-1} Z_j A_j b_j).$$

Then

$$\begin{aligned}
\langle v, w\rangle_q &= \sum_j m_j \left[\frac{1}{2}\mathrm{Tr}\left((b_j \wedge u_j)(A_j^{-1} Z_j A_j)^\dagger)\right) + (A_j^{-1} z^* - A_j^{-1} Z_j A_j b_j) \cdot u_j\right] \\
&= \sum_j m_j \left[u_j \cdot (A_j^{-1} Z_j A_j b_j) + (A_j^{-1} z^* - A_j^{-1} Z_j A_j b_j) \cdot u_j\right] \\
&= \sum_j m_j (A_j^{-1} z^*) \cdot u_j \\
&= z^* \cdot \sum_j m_j A_j u_j.
\end{aligned}$$

But $m_1 A_1 u_1 + m_2 A_2 u_2 = 0$ by the definition of $\mathfrak{C}_q$ so the two vectors are orthogonal. □

Proof of Corollary 2. Let C be a linear involution in $O(n-1)$. Then C is diagonalizable over $\mathbb{R}$ with eigenspace decomposition $\mathbb{R}^{n-1} = (C+I)\mathbb{R}^{n-1} \oplus (C-I)\mathbb{R}^{n-1}$ and eigenvalues $1, -1$ having multiplicities $n-k-1$ and k, respectively, where $k \in \{0, 1, \ldots, n-1\}$. Thus for each such C there is k and $A \in GL(n-1, \mathbb{R})$ such that $C = A^{-1} J_k A$ where J_k is the diagonal matrix $\mathrm{diag}(I_{n-k-1}, -I_k)$ and I_l indicating the $l \times l$ identity matrix. We can take A to be orthogonal. In fact, let $A = SU$ be the polar decomposition of A into a positive symmetric part $S = \sqrt{A^\dagger A}$ and orthogonal part U. The condition $C^\dagger C = I$ implies that S^2 and J_k commute, from which it follows that S^2, hence S, is also a block matrix with 0 on the off-diagonal blocks of size $k \times (n-k-1)$ and $(n-k-1) \times k$. Therefore, S commutes with J_k whence the claim. Thus the set of all orthogonal involutions in dimension $n-1$ is the disjoint union of the sets $\mathcal{J}_k = \{U^\dagger J_k U : U \in O(n-1)\}$. It is clear from this description that $\mathcal{J}_k$ is the homogeneous space $O(n-1)/L$, where L is the isotropy group of J_k. Equivalently, L is the subgroup of all U that commute with J_k, which is easily seen to be the product $O(n-k-1) \times O(k)$. □

The following proposition gives a concrete expression for the unit normal vector field.

Proposition 22. Let $\nu_j(b_j)$ denote the unit outward pointing normal vector to body B_j at the boundary point b_j. Then the unit normal vector to ∂M at q is given by

$$\mathbb{n}_q = (c_1(\mathcal{L}_1^{-1}(b_1 \wedge \nu_1(b_1)), \nu_1(b_1)), c_2(\mathcal{L}_2^{-1}(b_2 \wedge \nu_2(b_2)), \nu_2(b_2)))$$

where c_1, c_2 are defined up to a common sign by the equations $m_1 c_1 = m_2 c_2$ and

$$\sum_j c_j^2 m_j \left[1 + \frac{1}{2} Tr\left((b_j \wedge \nu_j(b_j))(\mathcal{L}_j^{-1}(b_j \wedge \nu_j(b_j)))^\dagger\right)\right] = 1.$$

Proof. The unit normal vector $\mathbb{n}_q$, being an element of $\mathfrak{C}_q$, can be written as

$$\mathbb{n}_q = ((\mathcal{L}_1^{-1}(b_1 \wedge u_1), u_1), (\mathcal{L}_2^{-1}(b_2 \wedge u_2), u_2))$$

for some $u_j \in \mathbb{R}^n$. Recall that a vector $v = ((Z_1, z_1), (Z_2, z_2))$ tangent to ∂M has the form

$$Z_j = \mathrm{Ad}_{\sigma_j}(Z - V_j)$$
$$z_j = \sigma_j z - \mathrm{Ad}_{\sigma_j}(Z - V_j) b_j - v_j$$

where V_j and v_j are related through the shape operators as discussed earlier and v_j is tangent to the boundary of body B_j at b_j. Let as before $\nu_j(b_j)$ denote the unit normal vector to body B_j at b_j. Using the explicit form of the Riemannian metric we obtain after straightforward computation that

$$0 = \langle \mathbb{n}_q, v \rangle_q = -\sum_j m_j v_j \cdot u_j.$$

This being true for all v_j implies that $u_j = c_j \nu_j(b_j)$. But $m_1 \sigma_1^{-1} u_1 + m_2 \sigma_2^{-1} u_2 = 0$ by the definition of $\mathfrak{C}_q$ and $\sigma_j^{-1} \nu_j(b_j) = -(-1)^j e_n$. Thus the first equation. The second equation corresponds to the condition $\|\mathbb{n}_q\|^2 = 1$. □

Chapter 4

The Dynamics of No-slip Billiards

In this chapter we investigate the dynamics of no-slip billiards in more detail. In the first section we return to the no-slip strip of Example 1, giving an alternate proof of boundedness and considering periodicity. Section 4.2 focuses on the wedge of Example 2, developing the notion of the axis of periodicity which lead to the proof of Theorem 3. In the final section we take a closer look at the no-slip circle and its double circle caustic. But first, we review the no-slip transformation, uniting the alternate descriptions given in the two preceding chapters.

If the unconstrained model for two rigid bodies in Chapter 3 is adapted to one fixed body and a non-intersecting but otherwise unconstrained disk of uniform mass distribution, then rewriting Equation 3.1.3 using a rotation matrix, the post-collision rotational and linear velocity (v_0^+, v^+) is the function of the pre-collision velocities (v_0^-, v^-) given by

$$\begin{aligned} v_0^+ &= -\frac{1}{3}v_0^- + \frac{2\sqrt{2}}{3}v \cdot (R_{\frac{\pi}{2}}\nu) \\ v^+ &= \left[\frac{2\sqrt{2}}{3}v_0^- + \frac{1}{3}v^- \cdot (R_{\frac{\pi}{2}}\nu)\right] R_{\frac{\pi}{2}}\nu - (v^- \cdot \nu)\nu, \end{aligned} \tag{4.0.1}$$

where in general we write R_θ for the matrix of counterclockwise rotation by angle θ and ν is the inward normal vector at the point of contact. For dimension two we may use coordinates $x = (x_0, x_1, x_2)^\dagger$, with planar position $(x_1, x_2) = (x, y)$ and normalized rotational position $x_0 = \frac{R}{\sqrt{2}}\theta$, where R is the radius of the disk. If the upward normal is in the direction x_2, then by 4.0.1 the pre- and post-collision velocities are related by $v^+ = Tv^-$ where $v = (\dot{x_0}, \dot{x_1}, \dot{x_2})^\dagger$

and the transformation matrix $T \in O(3)$ is

$$T = \begin{pmatrix} -\frac{1}{3} & \frac{2\sqrt{2}}{3} & 0 \\ \frac{2\sqrt{2}}{3} & \frac{1}{3} & 0 \\ 0 & 0 & -1 \end{pmatrix}, \tag{4.0.2}$$

identical to 2.1.8.

4.1 Parallel Boundaries

The following strengthening of this main result in [5] holds for unit velocity and fixed unit separation, letting x_1 be the parallel direction, so that the time between collisions will be $t = \frac{1}{|\dot{x}_2|}$, constant as $\dot{x}_2$ only changes sign at collisions.

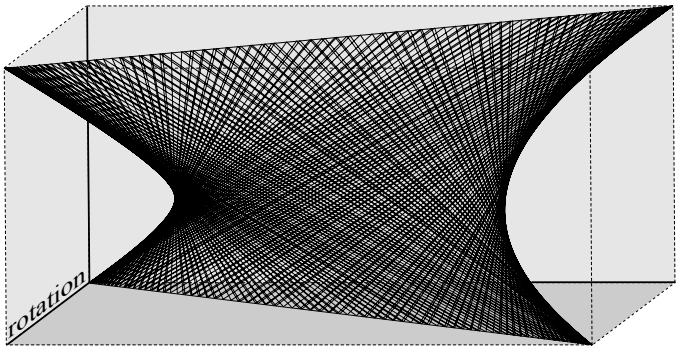

Figure 4.1: One orbit segment of the no-slip strip viewed with rotational position as the third dimension. Figure 1.2 is the projection of this one to the plane of the billiard table.

Proposition 23. Orbits of a two dimensional system of no-slip collisions with parallel boundaries have horizontal displacement no more than $\sqrt{\frac{3}{2}\left(\frac{1}{(\dot{x}_2)^2} - 1\right)}$.

Proof. Viewing the trajectories in three dimensions, we will show that for each boundary plane the set of collision points is contained in a line (Figure 4.1). Consider any two successive

collision points q^- and q^+ on the upper wall, choosing coordinates with the origin at the intermediate collision on the lower wall and letting $\dot{x} = (\dot{x}_0, \dot{x}_1, \dot{x}_2)$ be the velocity after the collision at q^-. Then $t = -\frac{1}{\dot{x}_2}$ and 2.1.8 gives

$$q^- = \left(\frac{\dot{x}_0}{\dot{x}_2}, \frac{\dot{x}_1}{\dot{x}_2}, 1\right)$$

and

$$q^+ = \left(\frac{\frac{1}{3}\dot{x}_0 - \frac{2}{3}\sqrt{2}\dot{x}_1}{\dot{x}_2}, \frac{-\frac{1}{3}\dot{x}_1 - \frac{2}{3}\sqrt{2}\dot{x}_0}{\dot{x}_2}, 1\right).$$

The slope in the x_1x_0 (horizontal-rotational) plane is

$$\frac{\Delta x_1}{\Delta x_0} = \frac{\frac{2}{3}\sqrt{2}\dot{x}_0 + \frac{4}{3}\dot{x}_1}{\frac{2}{3}\dot{x}_0 + \frac{2}{3}\sqrt{2}\dot{x}_1} = \sqrt{2},$$

independent of the initial conditions. Therefore, all of the collision points on the upper wall lie on a line of slope $\sqrt{2}$, and a similar argument on the lower wall shows that all collisions occur on a line of slope $-\sqrt{2}$.

As no velocity is exchanged between the horizontal-rotational component and the vertical component, the length of segments representing the projection of the orbits between successive collisions will be $\sqrt{\frac{1}{(\dot{x}_2)^2} - 1}$. Geometrically, the orbits are contained in an astroid (Figure 4.2) with the given bound, the maximal horizontal displacement being achieved when the (projection of) the trajectories is perpendicular to the contact line. □

Remark 1. The no-slip strip can never have more than three consecutive trajectory segments in the same horizontal direction: considering the trajectory segments in the upper, left, lower, and right quadrants of the x_0x_1 plane as delineated by the contact lines (as in Figure 4.2), no two consecutive trajectories both lie in the upper (or lower) quadrants. At most two consecutive trajectories may lie in the left (or right) quadrants.

In the case of three spacial dimensions, the argument of Proposition 23 may be generalized

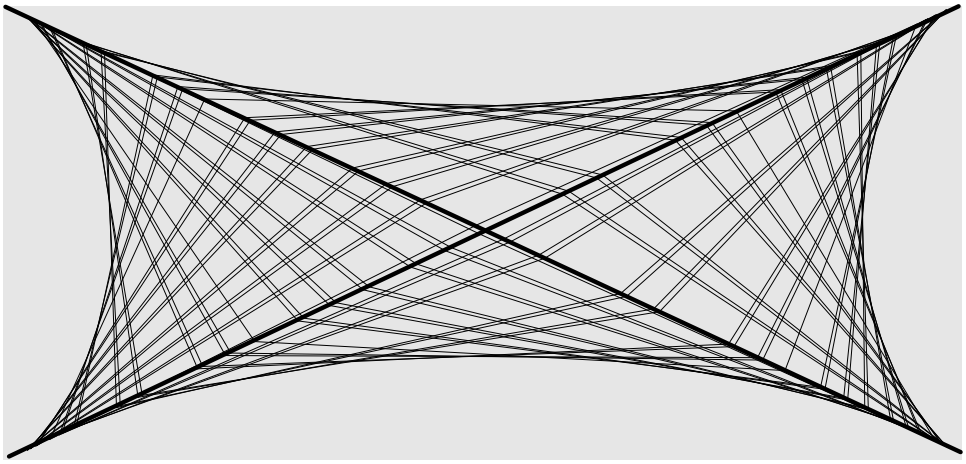

Figure 4.2: For the no-slip strip, the projections of the orbits onto the horizontal-rotational plane are contained in an astroid. Each orbit must be a segment of fixed length in this plane from one contact line to the other.

to two hyperplanes of the phase space in both of which the projected velocity is constant. The contact points occur on lines of slope $\sqrt{\frac{5}{2}}$ and similar bounds can be obtained in terms of the constant velocity in (or excluding) the hyperplanes, yielding the following result suggested by Example 6.

Proposition 24. Orbits of three dimensional no-slip collisions between parallel planes are bounded.

Next we consider the question of periodic orbits in the strip case. If the velocity is entirely vertical there will be a period two orbit, where the disk's velocity is vertical with no rotation. This is a simple example of an axis of periodicity. It is useful to consider the composition of transformations on the velocity through a complete cycle of two collisions. In the no-slip strip case, $v^+ = Sv^-$ where $S \in SO(3)$ is given by

$$S = (FT)^2 = \begin{pmatrix} -\frac{7}{9} & -\frac{4}{9}\sqrt{2} & 0 \\ \frac{4}{9}\sqrt{2} & -\frac{7}{9} & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{4.1.1}$$

with T as in equation 2.1.8 and F the appropriate frame adjustment.

The fact that there are no higher order periodic orbits for the strip will follow from considering iterations of S and the possible values of $\cos\theta$ when θ is a rational multiple of π. Suppose $\theta = \frac{p}{q}\pi$ for $p, q \in \mathbb{Z}$ and $\cos(\theta) \in \mathbb{Q}$. By Euler's identity

$$2\cos(\theta) = e^{-\frac{p}{q}\pi i} + e^{\frac{p}{q}\pi i}. \tag{4.1.2}$$

For any integers p, q, $e^{-\frac{p}{q}\pi i}$ and $e^{\frac{p}{q}\pi i}$ are roots of $x^q \pm 1$ and therefore are in the ring of algebraic integers. By 4.1.2 their sum $2\cos(\theta)$ is also an algebraic integer and can only be rational if it is an integer [26], namely -1, 0, or 1, proving the following lemma of Niven [22].

Lemma 25. If θ is a rational multiple of π, the only possible rational values of $\cos(\theta)$ are $0, \pm\frac{1}{2}$, and ± 1.

Proposition 26. The no-slip billiard system on the infinite strip has no periodic orbits besides the trivial period-two orbit.

Proof. Notice that $v = Sv$ only if the velocity is vertical, giving the trivial period two orbit. Suppose $\dot{x_0} \neq 0$ or $\dot{x_1} \neq 0$. A necessary condition for higher order periodicity is then that $v = S^n v$ for some $n > 1$. $S \in SO(3)$ is a rotation matrix and will have finite order precisely when it is a rotation through an angle which is a rational multiple of π. But here the angle is $\alpha = \cos^{-1}(\frac{7}{9})$ is not a rational multiple of π by Lemma 25. □

It follows from Proposition 26 that the bounds of Proposition 23 are optimal. If $\dot{x}_2 = 1$ then the bound is 0 which holds trivially for the simply periodic case. Otherwise, the orbit is not periodic and will come arbitrarily close to the geometric limit on the astroid by invariance of measure. Notice that for a certain initial velocity it is possible to have a ray-like orbit which achieves the limit on one side and then reverses direction. However, by the non-existence of higher order periodic orbits it cannot also achieve the limit on the other side.

The matrix S and corresponding rotation angle α in the proof of Proposition 26 will generalize. S has an eigenvalue of 1 corresponding to the eigenvector $(0,0,1)$, the axis of periodicity. Because the axis corresponds to a coordinate axis, S is easily identified with an $S' \in SO(2)$, which suggests a natural connection to the [5] proof in terms of rotation in the complex plane. For the wedge example, though, the axis will have a component in the rotational direction; however, the axis and rotation angle will still be fundamental.

Another approach to the no-slip strip which provides a closed form expression for the horizontal displacement and can be generalized to the case of the wedge in the next section is to view the velocity vector at collision points in coordinates (ψ, r, y) relative to the axis of periodicity, where y is the velocity component along the axis of periodicity–which is just the spacial axis perpendicular to the walls in this case–and ψ, r represent the component in the perpendicular plane in polar coordinates, taking ψ to be the angle relative to the positive horizontal direction x_1. In these coordinates, if $v^{(0)} = (\psi_0, r_0, y_0)$ then the no-slip transformation is given by

$$Tv^{(0)} = (\beta + \psi_0 + \pi, r_0, -y_0).$$

If we let $v^{(0)} = (0, r, \sqrt{1-r^2})$ for an arbitrary $0 < r < 1$, then

$$v^{(k)} = (r\sin(\psi + k(\beta+\pi)), r\cos(\psi + k(\beta+\pi)), \sqrt{1-r^2}).$$

The time between collisions is then (assuming unit separation) $\tau = \frac{r}{\sqrt{1-r^2}}$ and the horizontal displacement after $k+1$ collisions is given by $h_{k+1} = h_k + \tau r\cos(\psi + k(\beta+\pi))$, or

$$h_k = \frac{r}{\sqrt{1-r^2}} \sum_{j=0}^{k-1} \cos(\psi + j(\beta+\pi)).$$

By the Lagrange trigonometric identity we get the the closed form

$$h_k = \frac{\sqrt{3}r}{\sqrt{2}\sqrt{1-r^2}} \sin\left(\frac{k}{2}(\beta+\pi)\right)\cos\left(\psi + \frac{k-1}{2}(\beta+\pi) + \frac{\pi}{2}\right). \tag{4.1.3}$$

4.2 Open wedges

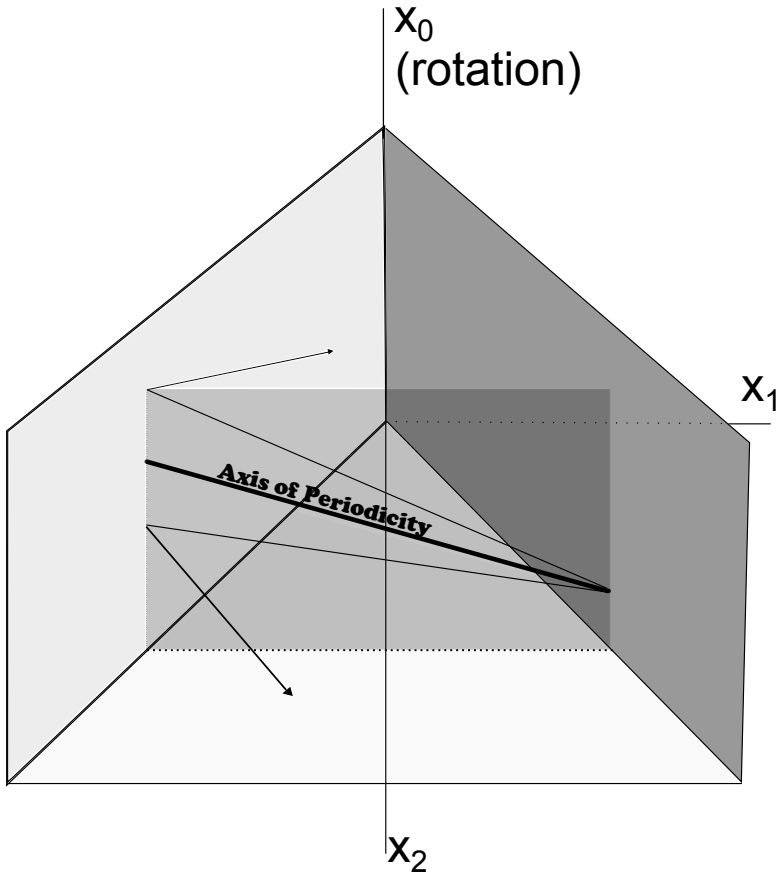

Figure 4.3: Adjusting the rotational velocity (the vertical component) appropriately will give a 2-periodic orbit.

In this section we return to Example 2, looking at no-slip billiards on an open wedge of angle θ, and prove Theorem 3. Let x_2 be the direction of the outward bisector, x_1 the perpendicular spatial direction, and x_0 the rotational position. We begin by considering an important type of periodic orbit which will always exist in wedge billiards.

Definition 11. A trajectory between two successive contact points in a no-slip billiard is an *axis of periodicity* if the direction is reversed at collisions, resulting in a 2-periodic orbit.

Informally, consider trajectories in the plane orthogonal to the x_2 (outward) direction. Due to the wedge angle, a trajectory with small rotational component, shown in the vertical direction in Figure 4.3, would reflect in an outward direction, while a large enough rotation would reflect inward towards the wedge vertex after a collision. As the transformation is continuous, there is a direction which will exactly reflect. By this process, given any two points in a two dimensional no-slip billiard with no obstructions there exists a rotational direction which will yield a single reflection. Notice, however, that for the wedge the change of frames between collisions is given by rotating the x_0x_1 plane around the x_2 axis, a symmetry which ensures subsequent collisions will remain along the axis. Figure 4.4 shows the case $\theta = \frac{\pi}{2}$.

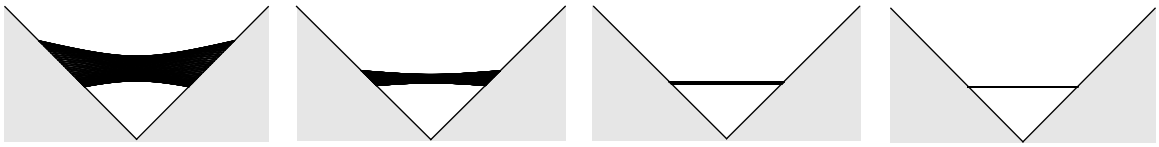

Figure 4.4: The trajectories for the wedge angle $\frac{\pi}{2}$ and initial velocities (-1,1,$\dot{x}_2$), satisfying Equation 4.2.1, with $\dot{x}_2$=0.25, 0.1, 0.02, and 0. As $\dot{x}_2$ approaches zero the orbit approaches the axis of periodicity.

Formally, the following holds:

Proposition 27. An orbit of an angle θ wedge billiard with initial velocity $(\dot{x}_0, \dot{x}_1, \dot{x}_2)$ will have period two if and only if

$$\frac{\dot{x_0}}{\dot{x_1}} = -\sqrt{2}\sin\frac{\theta}{2} \tag{4.2.1}$$

and $\dot{x_2} = 0$.

Proof. Consider $S_\theta \in SO(3)$ giving the transformation of the velocity after two collisions for a wedge of angle θ. Adjusting the normal for the angle, we have $S_\theta = (R'_{\frac{\theta}{2}} T R'_{-\frac{\theta}{2}})^2$, where R'_θ is

the frame adjusted rotation. Then $v = (-\sqrt{2}\sin\frac{\theta}{2}\dot{x_1}, \dot{x_1}, 0)$ is an eigenvector of S corresponding to the eigenvalue 1. Since the outward component $\dot{x}_2$ is zero, the position as well as the velocity is unchanged. □

The no-slip strip case might be thought of as the limit as θ approaches zero. (Note, that the two spatial axes are reversed here: 4.2.1 would imply $\dot{x}_0 = 0, \dot{x}_1 \neq 0$, giving the x_2 axis as the axis of periodicity.)

Turning to the question of higher periodicity for no-slip wedges, it is again useful to approach the question in terms of the velocity transformation S_θ. In order to have an orbit of period $2n$ it is necessary for the velocity to return to the initial velocity after n iterations, hence a necessary condition is $v = S_\theta^n v$. As S_θ simply rotates the velocity vector around the axis of periodicity, this condition will depend on the rotation angle α being a rational multiple of π. Notice that $w = (0, 0, 1)$ is orthogonal to the v given in Proposition 27 giving the axis of periodicity. Then $\cos(\alpha) = v \cdot S_\theta v$. A direct calculation gives the following relation between θ and α.

Proposition 28. The wedge angle θ and the rotational angle α of $S_\theta \in SO(3)$, the corresponding transformation of the velocity after one cycle of two no-slip collisions, are related by

$$\frac{32}{9}\cos^4\left(\frac{\theta}{2}\right) - \frac{16}{3}\cos^2\left(\frac{\theta}{2}\right) + 1 = \cos(\alpha).$$

Corollary 4. For any $n \geq 2$ there is an angle $\theta \in (0, \pi)$ such that all (nondegenerate) bounded orbits in an angle θ wedge have period $2n$.

Proof. By the proposition θ can be chosen so that the velocities are transformed by a rotation of $\alpha = \frac{\pi}{n}$ around the axis of periodicity in the half-sphere representation of the velocities, ensuring the necessary condition that the velocities are periodic. Preservation of measure

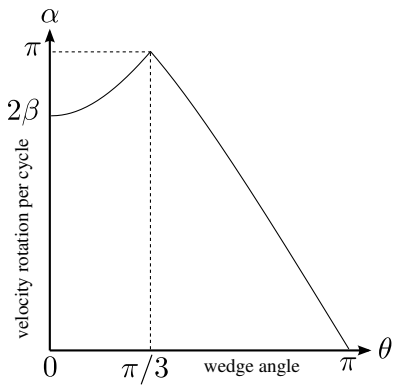

Figure 4.5: The velocity vector is rotated about the axis of periodicity by an angle α, determined by the wedge angle θ.

implies that this condition is also sufficient to ensure the orbit is periodic. (See Lemma 30 below.) Degenerate cases may occur with period less than $2n$, but a small perturbation of initial conditions will yield an orbit of maximal period. □

Corollary 5. The set of wedge angles θ giving periodic orbits is dense in $(0, \pi)$.

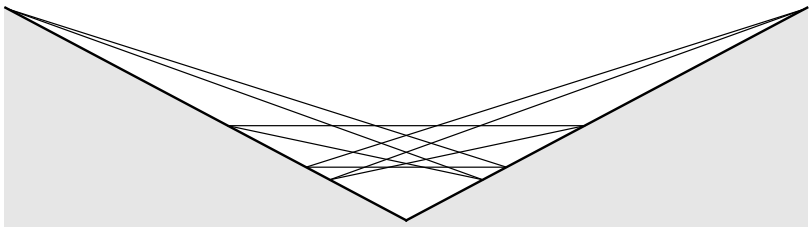

Figure 4.6: For any $n \geq 2$ a wedge angle θ can be chosen such that all bounded orbits are periodic of period $2n$. For $n = 4$, $\theta \approx 2.16598$ gives a (two-cycle) rotation angle $\alpha = \frac{\pi}{2}$ and period 8 trajectories.

Figure 4.6 shows the predicted period eight example. The degenerate cases occur where the spacial projection contains repeated segments; these are resolved when the rotational dimension is incorporated. This may happen in any higher order periodic case with appropriate initial conditions.

Recall that the strip produced bounded orbits for all initial conditions except for the horizontal trajectory. With the wedge there are many trajectories that escape, but the

analogous result below says that the velocity phase space may be similarly partitioned. By the assumption of unit velocity, the set of velocities may be identified with the sphere S^2, excluding the two $\dot{x}_0$ poles representing the cases in which the there is only rotational motion. By time reversibility antipodal points correspond to the same trajectories and may be identified. The existence of an axis of periodicity for all wedge angles θ implies the region of S^2 corresponding to bounded orbits is not empty. Numerically, one finds two antipodal non-escape regions with area a decreasing function of θ, approaching 0 as θ approaches $\frac{\pi}{2}$ and growing to 2π, an entire hemisphere, as θ approaches 0. The latter limit is consistent with the fact that almost all initial conditions yield bounded orbits in the no-slip strip.

To specify these regions formally, let E_θ^0 be the sector of the sphere corresponding to directions of direct escape from the θ wedge, along with the antipodal sector. Then for $S_\theta \in SO(3)$ we may define the escape region for θ as

$$E_\theta = \bigcup_{n\in\mathbb{Z}} S_\theta^n E_\theta^0,$$

and its complement E_θ^c will be the region of non-escape velocities.

Proposition 29. Every orbit with a non-escape velocity remains in the non-escape region.

Proof. First suppose θ is not periodic, that is, the corresponding angle of rotation for the transformation S_θ is not a rational multiple of π. Distance from the axis of periodicity is invariant under rotation by S_θ, so the closest point in E_θ^0 is rotated around a fixed spherical circle whose center is the axis point. Every point outside of it eventually is mapped to the escape sector, while no points inside are. Hence the escape and non-escape regions partition the sphere, except for a circular boundary. See Figure 4.7.

If θ is periodic of period $2n$, then the non-escape region E_θ^c is a curvilinear regular polygon, an n-gon if n us odd and a 2n-gon in n is even, which remains fixed with sides permuted under S_θ.

□

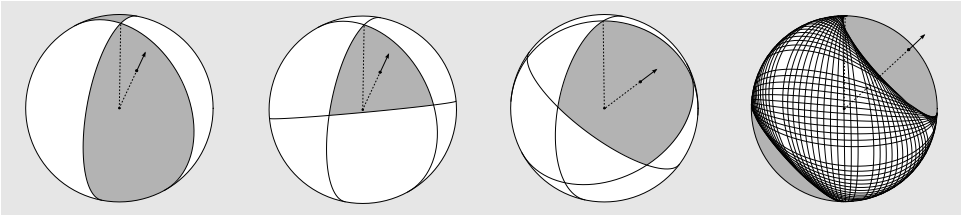

Figure 4.7: The velocity space may be represented as a sphere for wedge systems. The velocity transformation rotates the escape sector (light area, left) around the axis of periodicity. The dark region for the right three spheres is the non-escape region, with polygonal boundary for periodic θ (left middle, right middle) and caustic circle for non-periodic θ (right).

We can now prove the following lemma which was used above in establishing periodicity.

Lemma 30. For a given orbit, if two collision points on the same edge of a no-slip wedge have identical velocities then their positions also correspond and the orbit is periodic.

Proof. Suppose an orbit has a velocity vector repeated at distinct positions, and consider the trajectory of k collisions between the two points oriented in the direction from the outer point to the inner point. The velocity pattern will then repeat the same cycle of k trajectories as the orbit continues inward indefinitely. Consider the action of k iterations of the billiard map on the subset of phase space consisting of the product of the non-escape velocities and a small rectangle in the rotational and spacial dimensions. The velocities are invariant while the rectangle contracts exponentially in the spacial dimensions and expands at most linearly in the rotational dimension, which violates preservation of measure.

□

Using the approach of the alternative proof of the boundedness for the no-slip strip at the end of the last section, it is possible to describe the displacement in the direction of the wedge bisector, of interest in establishing that the orbits are bounded, by a product. Letting

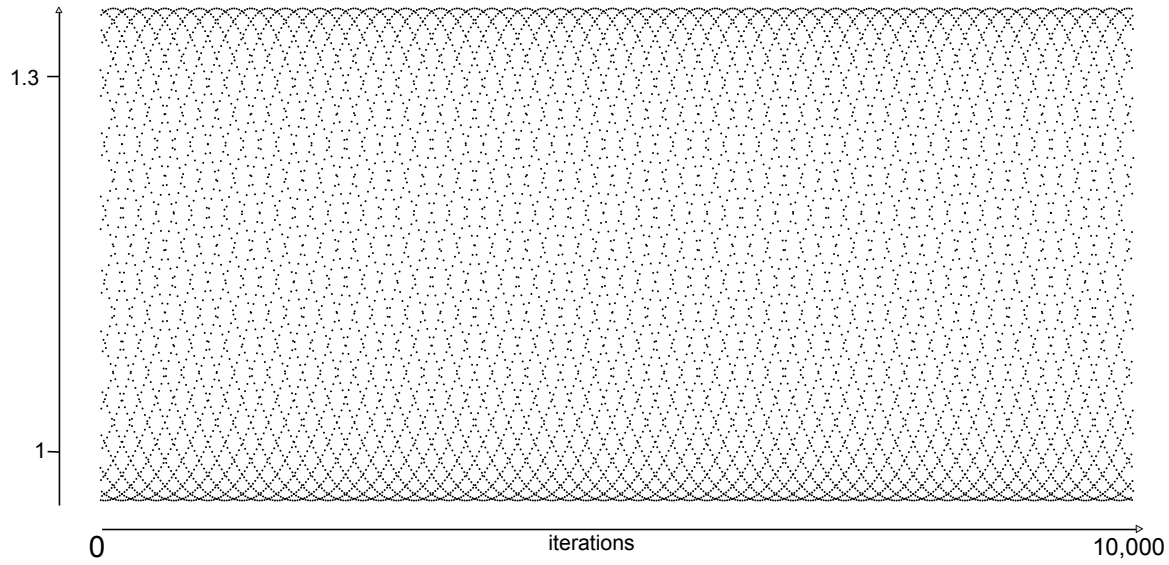

Figure 4.8: Graph of the wedge displacement for $\theta = \frac{\pi}{2}$ and $r = 0.1$ using the product formula.

$q^{(n+1)}$ be the position at the $n+1$ collision, we have

$$q^{(n+1)} = \prod_{k=0}^{n} \left(1 + \frac{2}{\frac{\cos\frac{\theta}{2}}{\sqrt{2-\cos\theta}} \left(\sqrt{\frac{1}{r^2} - 1} \csc\frac{\theta}{2} \sec(n(\alpha_\theta + \pi)) + \sqrt{2} \tan(n(\alpha_\theta + \pi)) \right) - 1} \right), \quad (4.2.2)$$

where θ is the wedge angle and α_θ the corresponding rotation angle, with r the magnitude of the velocity vector in the direction orthogonal to the axis of periodicity, which remains constant at collisions. The corresponding closed form is not yet known, but numerically it appears bounded for most values of r and θ.

The following theorem summarizes the main results of this section.

Theorem 3. For a no-slip billiard wedge of angle $\theta \in (0, \pi)$, let x_0 be the rotational axis, x_2 the direction of the wedge bisector, and x_1 the perpendicular spacial direction.

i There exists a periodic axis, a direction in which all trajectories are periodic. Specifically,

for velocity $(\dot{x_0}, \dot{x_1}, \dot{x_2})$, the orbit will be periodic whenever

$$\frac{\dot{x_0}}{\dot{x_1}} = -\sqrt{2}\sin\frac{\theta}{2}.$$

ii For any $n \in \mathbb{Z}^+$, wedge angle θ_n can be chosen so that all non-escape velocities yield $2n$-periodic orbits. Furthermore, the set of all such θ_n is dense in $(0, \pi)$.

iii The angle ψ between the velocity and the axis of periodicity is invariant throughout an orbit, remaining unchanged after collisions.

4.3 Circles

In this section we give a characterization of the orbits of no-slip circular billiards. As noted in [29, 13], such systems generally have double circular caustics.

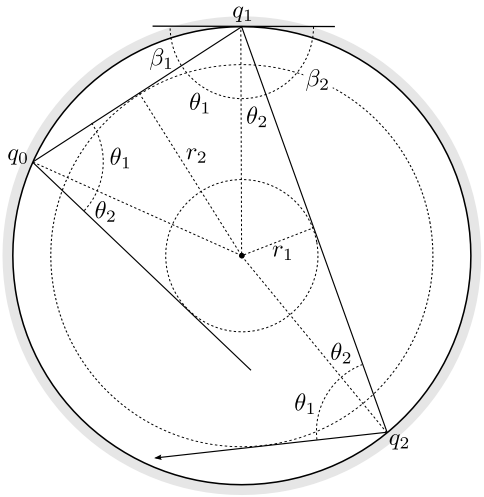

Figure 4.9: No-slip collisions result in alternating incident angles for circular billiards, with trajectories tangent to alternating circular caustics.

Proposition 31. For a billiard system with circular table of radius r and no-slip collisions, the projections of trajectories from the 3-dimensional angle-position space to the disk in

position plane have the property that the vertex angle at each collision is a constant of motion. Moreover, for each projected trajectory γ, there exists a pair of concentric circles of radius less than r that are touched tangentially and alternately by the sequence of line segments of γ at the middle point of these segments.

Implicit in the alternating caustics is the fact that the incoming angle relative to the tangent will alternate between angles β_1 and β_2 with successive collisions, or equivalently the internal angles between the trajectories and a radius will alternate between the complementary angles θ_1 and θ_2. (See Figure 4.9.)

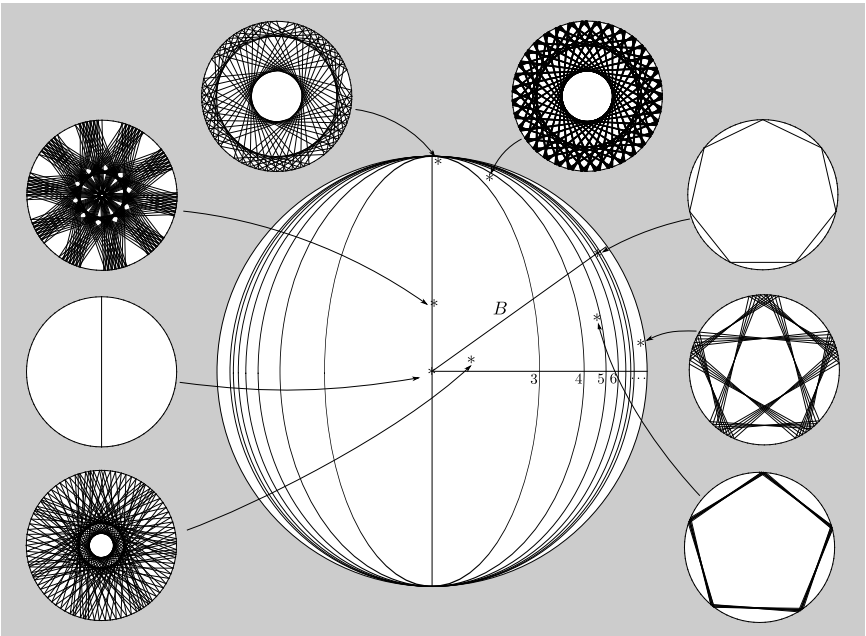

Figure 4.10: Projection of velocity phase space in rotational and tangential direction, with examples. The two caustics converge along B yielding orbits identical to those of standard billiards.

Consider three successive collisions, q_0, q_1, and q_2, and choose coordinates with q_1 at the origin, with rotational direction x_0, x_1 in the tangent direction, and x_2 in the radial direction. Let $v = (\dot{x}_0, \dot{x}_1, \dot{x}_2)$ be the velocity after the collision at q_0. All circular billiards up to a rotation or reflection may be considered by varying v and changing q_0 and q_2 accordingly, leaving q_1 fixed, hence requiring $\dot{x}_2 \neq 0$. Then the orbit is determined by the choice of v, and

in particular we can express the two caustic radii in terms of the components v_i. Immediately we have $r_1 = \sin\theta_1$ and $\tan\theta_1 = \frac{\dot{x}_1}{\dot{x}_2}$, and using the post-collision velocity Tv we also have $r_2 = \sin\theta_2$ and $\tan\theta_2 = \left(\frac{2\sqrt{2}}{3}\dot{x}_0 + \frac{1}{3}\dot{x}_1\right)/(-\dot{x}_2)$. Hence

$$r_1 = \frac{\dot{x}_1}{\sqrt{\dot{x}_1^2 + \dot{x}_2^2}} \tag{4.3.1}$$

and

$$r_2 = \frac{\frac{2\sqrt{2}}{3}\dot{x}_0 + \frac{1}{3}\dot{x}_1}{\sqrt{(\frac{2\sqrt{2}}{3}\dot{x}_0 + \frac{1}{3}\dot{x}_1)^2 + \dot{x}_2^2}} \tag{4.3.2}$$

Notice that in the above equations requiring the numerators to be equal guarantees the denominators are equal, thus we may ensure that $r_1 = r_2$ (and $\theta_1 = \theta_2$) by choosing v with $\dot{x}_1 = \sqrt{2}\dot{x}_0$. If we further require $\dot{x}_2 = \tan\left(\frac{\pi}{n}\right)\dot{x}_1$ the orbit will be a regular n-gon. These two requirements correspond respectively to segment B and the numbered elliptic curves in Figure 4.10, with regular n-gons at the intersections. Additionally, the family of orbits with coinciding caustics $r_1 = r_2$ are those in which there is no change in rotational velocity, a family corresponding precisely to the orbits of standard billiards. For standard billiards on a circle, or more generally on a smooth convex boundary, there is a countably infinite family of n-periodic billiards for any $n \geq 2$. (See [27].) For no-slip billiards there is a continuum of initial conditions that give periodic orbits for a given n.

Remark 2. For any boundary arc of angle greater than $\frac{\pi}{2}$, there is a positive measure region of phase for which the double caustics persist.

Chapter 5

Ergodic No-Slip Billiards

In this final chapter we look at the ergodic theory of no-slip billiards. The necessary prerequisite of determining that the no-slip map is invariant (at least in dimension two) is dealt with in Section 5.1. In Section 5.2 we discuss ergodic candidates.

5.1 The invariance of the no-slip map

A fundamental property of the dynamics of standard billiard systems is the existence of a canonical invariant measure on constant energy surfaces, sometimes referred to as the Liouville measure. We give here a sufficient condition for the same measure to be invariant under non-standard collisions.

Let S denote the boundary of the configuration manifold of the two-bodies system. As before, we assume that S is smooth. We fix a value $\mathcal{E}$ of the kinetic energy and denote

$$N^{\mathcal{E}} = \left\{ (q, v) \in TM : q \in S, \frac{1}{2}\|v\|^2 = \mathcal{E} \right\}.$$

Define the *contact* form θ on TM to be the 1-form such that $\theta_v(\xi) = \langle v, d\pi_v \xi \rangle_q$, where π is the base point projection from TM to M (and we indicate the element of TM by v rather than (q, v) in subscripts). It is well-known that $d\theta$ defines a symplectic form on TM. It can also be shown that the restriction of $d\theta$ to $N^{\mathcal{E}}$ defines a symplectic form on $N^{\mathcal{E}} \setminus TS$. (See, for example, [11].)

The *billiard map* T on $N^{\mathcal{E}}$ associates the post-collision state of the system at the time of a collision to the post-collision state at the next collision. There are well-known issues about

this map, even for standard billiards in dimension 2, that make the precise specification of its domain difficult to describe. See, for example, [9]. Here we assume that the domain of T consists of a "large" open set of full Lebesgue in $N^{\mathcal{E}}$ and omit any further reference to it since this issue of domains is not specific to our rough billiards. The next result is shown by a local argument and considerations of domain do not play a role.

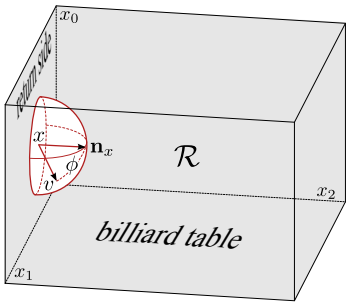

Figure 5.1: Angle-position parallelepiped $\mathcal{R}$ for rectangular billiard table and disc. The canonical invariant billiard measure on a constant energy hypersurface is, up to multiplicative constant, the product of the Euclidean area measure on the boundary of $\mathcal{R}$ and the measure on the hemisphere of velocity directions given by $\cos\phi\, dA$ where dA is the Euclidean area measure on the hemisphere. If billiard trajectories are initiated on the side $x_2 = 0$ with random initial condition given by the just described measure, return trajectories will have the same distribution. See Figure 5.2.

Theorem 4. Suppose that the field of collision maps $q \in S \mapsto \mathcal{C}_q$ is piecewise smooth and parallel (where it is smooth) with respect to the Levi-Civita connection associated to the kinetic energy Riemannian metric. Let $\Omega = d\theta \wedge \cdots \wedge d\theta$ be the form (of degree $2n-2$, where n is the dimension of the ambient Euclidean space) derived from the canonical symplectic form $d\theta$ on $N^{\mathcal{E}} \smallsetminus TS$. Then Ω is, up to sign, invariant under the billiard map.

The theorem will be proved in 5.1.1.

The boundary of M is a flat surface with the Euclidean metric and the vectors $e_0, J\nu(b)$ shown on the right-hand side of Figure 1.4 constitute a parallel frame. The orthogonal line distributions $\mathfrak{C}_q$ and $\mathfrak{S}_q$ are also parallel as the angle between each of them and e_0 is constant. But these are the eigenspaces of $\mathcal{C}_q$ for the eigenvalues -1 and 1, respectively. It follows

that the field of rough collision maps is parallel, yielding the following essential result for considering questions of ergodicity in the two dimensional case.

Theorem 5. The canonical billiard measure on the boundary of the phase space of a planar no-slip billiard is invariant under the billiard map.

We illustrate Theorem 4 with a numerical observation concerning the motion of a disc in a rectangular table with rough collisions. The geometric set-up is shown in Figure 5.1. The configuration manifold in this case is a parallelepiped $\mathcal{R}$ in dimension 3 and the canonical billiard measure on the manifold N of unit length vectors with base points on the boundary of $\mathcal{R}$ has density proportional to $\rho(v) = v \cdot \mathbb{n}_q = \cos\phi$, $0 \leq \phi \leq \pi/2$, with respect to the Riemannian volume measure on N, where ϕ is the angle the vector v makes with the normal vector to the boundary.

It can be shown that Theorem 4 applies to this case. As an experiment to illustrate invariance of the billiard measure for rough collisions we sample initial conditions on the face $x_2 = 0$ with the uniform distribution for the (x_0, x_1) positions and initial velocity v having probability density proportional to $\cos\phi$ relative to the uniform probability on the unit hemisphere. If the return states to the face $x_2 = 0$ are distributed according to the same measure, then the angle ϕ for the return velocity must be distributed relative to Lebesgue measure on $[0, \pi/2]$ with density $\sin(2\phi)$, which is the marginal density function for the angle distribution with respect to the Lebesgue measure $d\phi$ on $[0, \pi/2]$, under the assumption that the billiard measure is invariant.

This is indeed the case as shown in Figure 5.2. A large number (10^5) of initial conditions starting from one side of the rectangle are sampled from the normalized billiard measure restricted to that side. For each trajectory, the return state to that side is computed and the angle relative to the normal (to the side of the angle-position parallelepiped corresponding to that side of the rectangle) is recorded. The distribution of values is shown in the above

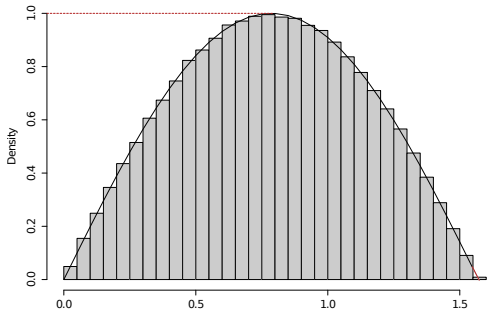

Figure 5.2: Experiment to illustrate invariance of the billiard measure for a rectangular table.

histogram. The superimposed line is the graph of $\sin(2\phi)$.

5.1.1 Proof of Theorem 4

Define the one-form θ on TM from the kinetic energy Riemannian metric on M so that $\theta(\xi) = \langle v, d\pi_v \xi \rangle$ for each $\xi \in T_{q,v}(TM)$, where $d\pi_v$ is the map induced on the tangent space at (q, v) of the base-point projection map $\pi : TM \to M$. We briefly recall the definition of the vertical and horizontal subbundles E^{V} and E^{H} of $T(TM)$. For simplicity of notation we denote points in TM by v rather than (q, v). Then the fiber E^{V}_v above v is the kernel of $d\pi_v$ and E^{H}_v is the kernel of the *connection map* $K_v : T_v(TM) \to T_q M$, defined as follows: if $\xi = w'(0)$ where $w(t)$ is a curve through v representing ξ, then $K_v \xi = \frac{\nabla}{dt}\big|_{t=0} w(t)$. If now X and Y are vector fields on TM, then

$$d\theta_v(X, Y) = \langle K_v X, d\pi_v Y \rangle - \langle K_v Y, d\pi_v X \rangle. \tag{5.1.1}$$

See [11] for more details.

Now let S denote the boundary of M, N the pull-back to S of the tangent bundle TM under the inclusion map and for each value $\mathcal{E} > 0$ define

$$N^{\mathcal{E}} := \left\{ (q, v) \in N : \frac{1}{2} \|v\|_q^2 = \mathcal{E} \right\}.$$

So $N^{\mathcal{E}}$ is a level set of the kinetic energy function. It is shown in [11], that the pull-back of $d\theta$ to $N^{\mathcal{E}}$ under the inclusion map is non-degenerate on $N^{\mathcal{E}} \setminus TS$, and so it defines there a symplectic form. If the ambient space of the system is $\mathbb{R}^n$ then $N^{\mathcal{E}}$ has dimension $2n-2$. The canonical billiard measure is now the measure associated to the $(2n-2)$-form $\Omega = (d\theta)^{n-1}$ pulled-back to $N^{\mathcal{E}} \setminus TS$.

The smooth field $q \mapsto \mathcal{C}_q$ of collision maps defines a smooth map (away from singularities) $\mathcal{C} : N^{\mathcal{E}} \to N^{\mathcal{E}}$. The pull-back of θ under this map is easily shown to be

$$(\mathcal{C}^*\theta)_v(\xi) = \langle v, \mathcal{C}_q d\pi_v \xi \rangle_q.$$

Note that $d\pi_v \xi \in T_q S$ whenever $\xi \in T_v N$. Define the projections $\Pi_q^{\pm}$ from $T_q S$ to the eigenspaces of $\mathcal{C}_q$ associated to eigenvalues ± 1. The assumption that $\mathcal{C}$ is parallel is equivalent to one of these projections (equivalently, both) being parallel. Now define $\theta_v^{\pm}(\xi) = \langle v, \Pi_q^{\pm} d\pi_v \xi \rangle$, so that $\theta = \theta^+ + \theta^-$ and $\mathcal{C}^*\theta = \theta^+ - \theta^-$. Consequently,

$$\mathcal{C}^* d\theta = d\theta^+ - d\theta^-.$$

The projections $\Pi^{\pm}$ can also be defined on $TN^{\mathcal{E}}$ by requiring

$$\Pi_q^{\pm} d\pi_v = d\pi_v \Pi_v^{\pm}, \quad K_v \Pi_v^{\pm} = \Pi_q^{\pm} K_v.$$

Using these maps we define 2-forms $\omega^{\pm}$ by $\omega_v^{\pm}(\xi, \eta) := d\theta_v(\Pi_v^{\pm}\xi, \Pi_v^{\pm}\eta)$. We now wish to relate $\omega^{\pm}$ and $d\theta^{\pm}$.

First define a tensor field $\vartheta^{\pm}$ on S such that for $u, v \in T_q S$ and any vector fields X, Y on S such that $X_q = u$ and $Y_q = v$, we have

$$\vartheta_q^{\pm}(u, v) := (\nabla_u \Pi^{\pm})Y - (\nabla_v \Pi^{\pm})X.$$

It is not difficult to verify that this is indeed a tensor field and the definition does not depend on the extensions X, Y of u, v. Furthermore, $\vartheta^{\pm}$ vanishes under the conditions of Theorem 4. A straightforward calculation now shows that

$$d\theta_v^{\pm}(\xi,\eta) = \omega_v^{\pm}(\xi,\eta) + \langle v, \vartheta^{\pm}(d\pi_v\xi, d\pi_v\eta)\rangle.$$

Therefore, $d\theta^{\pm} = \omega^{\pm}$ when the field of collision maps is parallel. Moreover, $d\theta = \omega^{+} + \omega^{-}$ and $\mathcal{C}^{*}\omega^{\pm} = \pm\omega^{\pm}$. It is now easy to check that

$$(d\theta)^{n-1} = \pm(\omega^{+})^{n_{+}} \wedge (\omega^{-})^{n_{-}}$$

where $n_{\pm}$ are the dimensions of the eigenspaces of $\mathcal{C}_q$ associated to eigenvalue ± 1, and we finally obtain $\mathcal{C}^{*}(d\theta)^{n-1} = \pm(d\theta)^{n-1}$. Therefore, the measure induced by $(d\theta)^{n-1}$ is invariant under $\mathcal{C}$.

5.2 Ergodic considerations

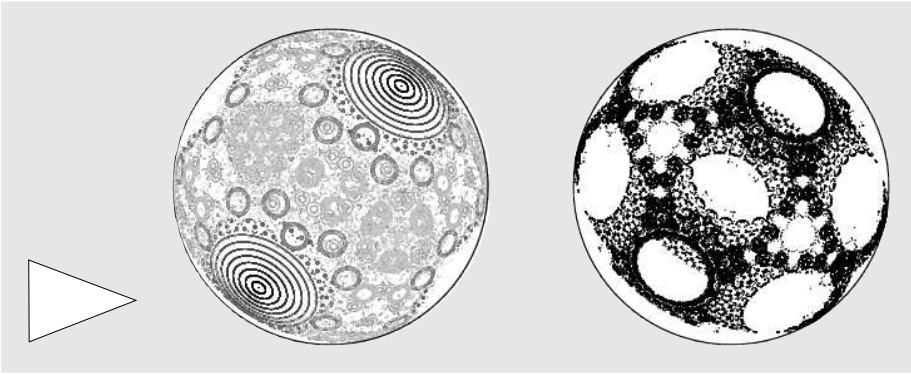

Figure 5.3: (Left) A sampling of phase space of the no-slip isosceles triangle. The axis of periodicity yields the two prominent pairs of concentric circles in the upper right and lower left, with many neighborhoods of higher order periodic points creating more complex symmetries. (Right) A single orbit near a high order periodic point.

We begin by looking at (the projection of) the velocity phase portraits of polygons. Wedge examples support Conjecture 1 and suggest that polygon portraits should exhibit a pervasive elliptic structure. Indeed, that is the case, as in the velocity phase portrait of an isosceles triangle no-slip billiard is given in Figure 5.3. Note that the overlapping orbits, a feature absent in standard billiards, are the result of the projection.

This pattern is typical of polygons, with most orbits appearing as closed curves around periodic points, the number of components matching the period of the central point. Numerical evidence suggests that for no-slip polygons, not only is the no-slip map not ergodic, but there are no positive measure ergodic components and extremely intricate orbits (as in Figure 5.3, right) are common. Experiments using a range of polygons and a range of variation in curvature have not produced examples free from evidence of non-ergodic behavior. Figure 5.4 gives the phase portraits for three small but increasing C^2 perturbations of the no-slip regular pentagon, demonstrating the persistence of the dynamics for small curvature.

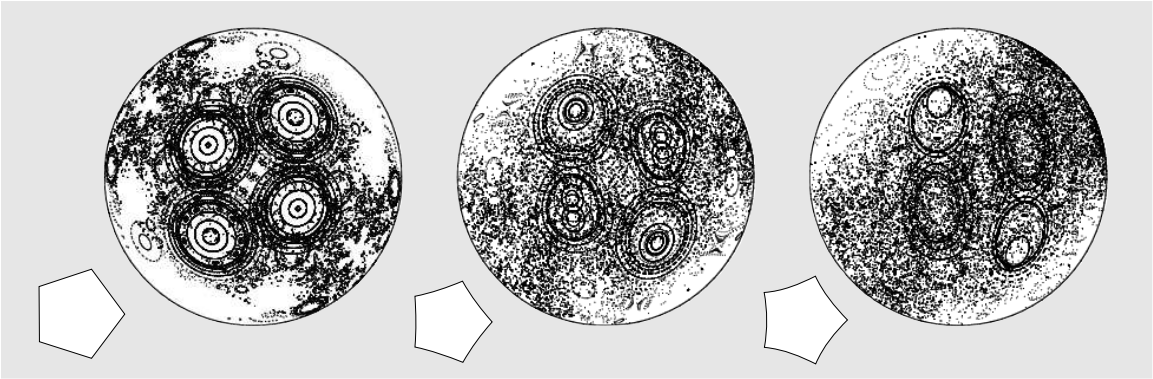

Figure 5.4: For small perturbations of the pentagon, the phase portraits suggest the non-ergodic dynamics persist.

Standard billiards with entirely concave boundaries are dispersing, with wave fronts expanding at every collision. In contrast, the chaotic behavior of the Bunimovich stadium depends upon a defocusing mechanism: convex boundaries create a focusing front, but under circumstances in which the front passes through a focusing point and sees a net expansion

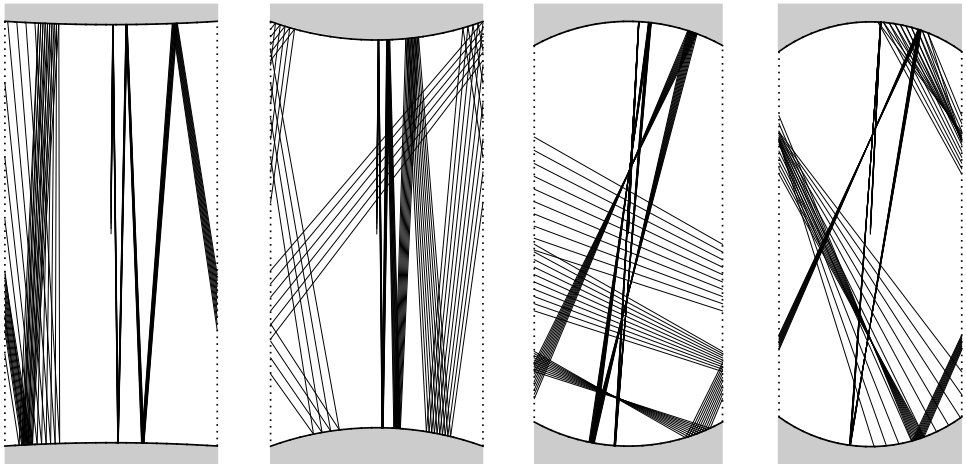

Figure 5.5: Straight edges are identified, so that collisions only occur at the curved boundaries. Left and middle left: examples of dispersion for standard and no-slip billiards. Middle right and right: examples of focusing.

before colliding at the next boundary. Looking at billiards with left and right edges identified and concave (or convex) upper and lower boundaries, both mechanisms appear to arise in the no-slip case as well. In Figure 5.5, we started with flat boundaries for both the standard and no-slip model, then gradually increased (decreased) the curvature until arriving at the examples above, in which dispersion (focusing) became evident. To trigger a comparable level of dispersion after the same number of collisions for an equally spaced wave front, a greater curvature is required for the no-slip case. Notably, however, the curvature required to produce focusing was similar, a phenomenon show rigorously in [29]. As the no-slip examples are projections, with independent expansion or contraction arising in the rotational dimension, the behavior may be more nuanced.

In spite of the apparent defocusing, the stadium has a bounded positive measure component phase space and is not ergodic. Similarly, other known standard billiard examples fail to be ergodic in the no-slip case: the 'mushroom' billiard [7] has a similar bounded component in the stem, the 'flower' billiard fails to be ergodic by Remark 2, while 'pocket' billiards are not ergodic for both of the above reasons. In these cases the no-slip phase portraits show a

mixture of structured orbits mixed surrounded by potential chaotic seas (Figure 5.6). The 'moon' billiard, recently shown to exhibit ergodic behavior for some parameters [12], does not appear to be ergodic in the no-slip case. Along with the mushroom billiard, its phase portrait exhibits segments along which orbits linger before moving to apparent chaotic regions. This behavior is qualitatively similar to the known *marginally unstable periodic orbits* of standard billiards which both the moon [12] and mushroom [1, 15] are known to exhibit.

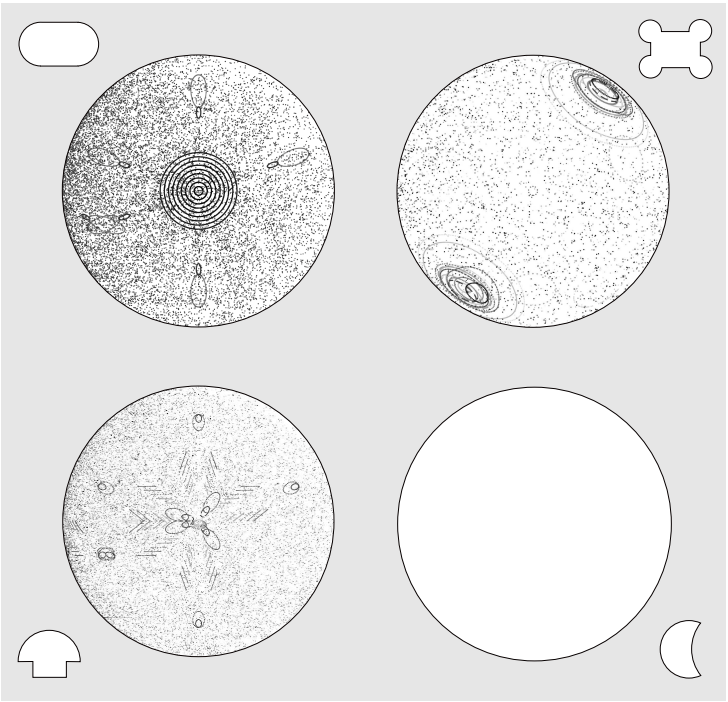

Figure 5.6: Velocity phase portraits of the stadium (upper left), the mushroom (lower left), a pocketed rectangle (upper right), and a moon billiard (lower right) exhibit closed orbits from bounded regions, but possibly large ergodic components as well.

Looking at the differential of the no-slip map in the case of 2-periodic orbits between collisions at boundary points with parallel tangents, Wojtjowski [29] showed ellipticity and linear stability for a sufficiently small product of curvature and orbit length relative to the moment of inertia. Applying this result to a Sinai billiard with a single dispersing disk on a torus, we may increase the curvature by decreasing the radius. By the formula in [29] applied

to this model, the breakdown in ellipticity should occur at the point when the radius becomes smaller than $\frac{2}{3}$. Looking at a small perturbation of the horizontal orbit, one finds that indeed the dispersion occurs rapidly once the threshold is passed.

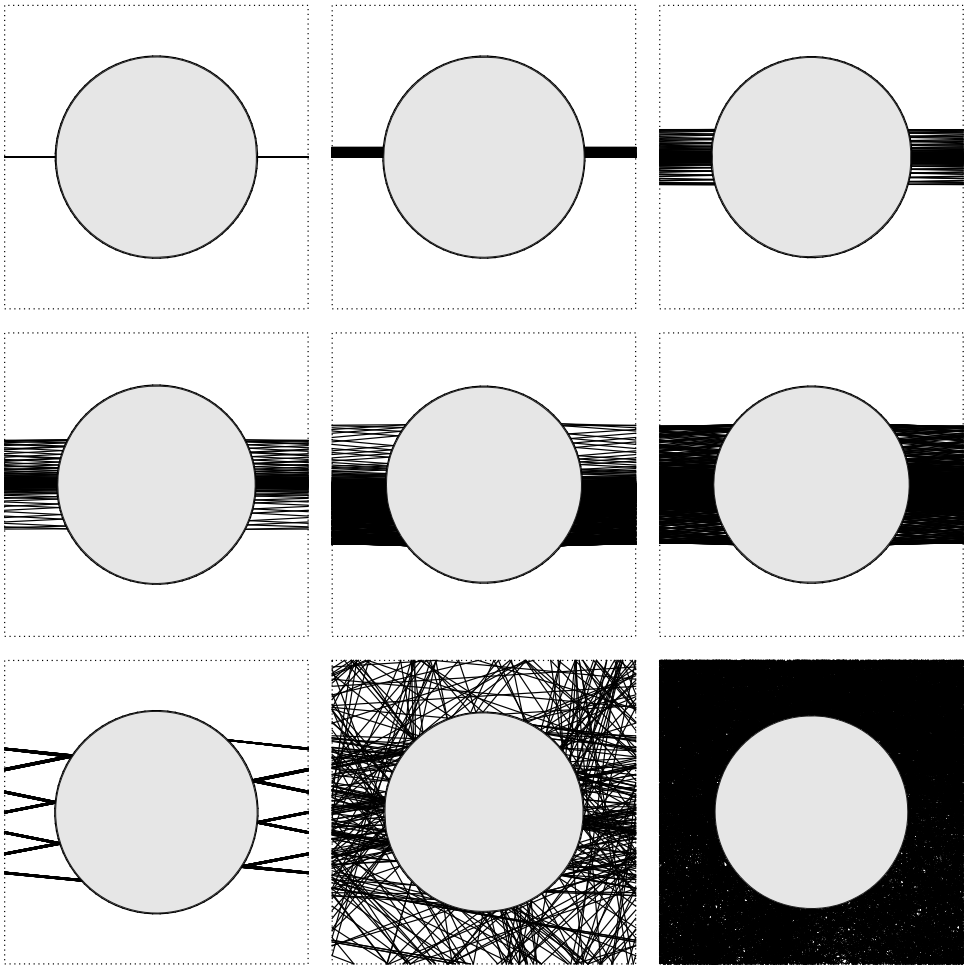

Figure 5.7: The orbit in the upper left corner is stable, with radius very slightly larger than $\frac{2}{3}$. The radius is increased slightly until in the bottom right (r = .664) the orbit is no longer stable and dispersion occurs.

Unlike the case of standard billiards, periodic points for this dispersing billiard may also appear for horizontal trajectories colliding at any point on the disk, with the appropriate rotational velocity to align with the axis of periodicity. Figure 5.8 shows the result of the corresponding experiment, which suggests that a similar dispersion occurs.

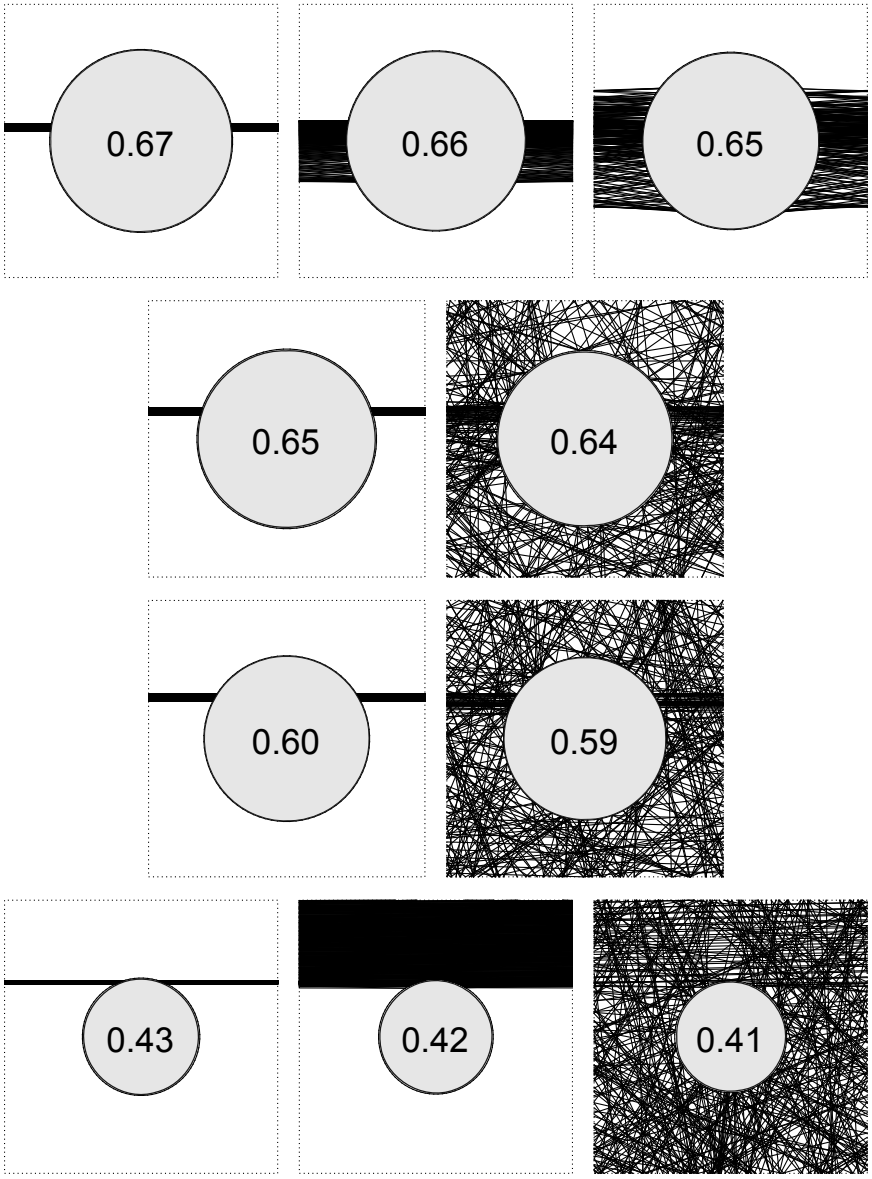

Figure 5.8: By row, trajectories at height 0.1, 0.2, 0.3, and 0.4, near an axis of periodicity. Numbers give the radius of the dispersing disk. More curvature is required than for the parallel case, but dispersion appears to occur. Note that the r=0.42 example (bottom middle) will continue until it reaches the bottom of the disk and then disperse, as the rotation will be reversed.

Bibliography

[1] E. G. Altmann, T. Friedrich, A. E. Motter, H. Kantz, A. Richter, *Prevalence of marginally unstable periodic orbits in chaotic billiards*. Phys. Rev. E 77, 016205 (2008).

[2] V. Arnold, *Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits*. Annales de l'institue Fourier, tome 16, n1 (1966), 319-361.

[3] A.M. Bloch, *Nonholonomic Mechanics and Control*, Springer, 2003.

[4] B. Hasselblatt and A. Katok, eds. Handbook of Dynamical Systems, V.1A, Elsevier 2002.

[5] D.S. Broomhead, E. Gutkin, *The dynamics of billiards with no-slip collisions*. Physica D 67 (1993) 188-197.

[6] L.A. Bunimovich, *The ergodic properties of certain billiards*. Funkt. Anal. Prilozh. 8 (1974) 73-74.

[7] L. A. Bunimovich, *Mushrooms and other billiards with divided phase space*. Chaos, 11 (2001) 802-808.

[8] L. A. Bunimovich *Dynamical billiards*. Scholarpedia, 2(8):1813 (2007).

[9] N. Chernov, R. Markarian, *Chaotic billiards*. Mathematical Surveys and Monographs, V. 127, American Mathematical Society, 2006.

[10] N. Chernov, R. Markarian. *Introduction to the Ergodic Theory of Chaotic Billiards*. http://people.cas.uab.edu/ mosya/papers/rbook.pdf.

[11] S. Cook, R. Feres, *Random billiards with wall temperature and associated Markov chains*. Nonlinearity 25 (2012) 2503-2541.

[12] M. F. Correia, H. K. Zhang, *Stability and ergodicity of moon billiards*. Chaos, 25 083110 (2015).

[13] C. Cox, R. Feres, W. Ward, *Differential geometry of rigid bodies collisions and non-standard billiards*. (arXiv:1501.06536)

[14] R. Cross, *Measurements of the horizontal coefficient of restitution for a superball and a tennis ball*. American Journal of Physics, (5) 70 (2002), 482-489.

[15] C. Dettmann, O. Georgiou, *Open mushrooms: stickiness revisited*. Journal of Physics A: Mathematical and Theoretical, vol 44, (2011).

[16] R. L. Garwin, *Kinematics of an ultraelastic rough ball*. American Journal of Physics, (1) 37 (1969), 88-92.

[17] J. Hadamard, Les surfaces à courbures opposées et leurs lignes géodésiques. J. Math. Pures et Appl. 4: 27–73, (1898).

[18] S. Kerckhoff, H. Masur, J. Smillie, *Ergodicity of billiard flows and quadratic differentials*. Ann. of Math. (2) 124 (1986), no. 2, 293-311.

[19] V. Kozlov, D. Teshchëv, *Billiards: a genetic introduction to the dynamics of systems with impacts*, Translations of Mathematical Monographs, V.89, American Mathematical Society, 1991.

[20] R. D. Lorenz, *Spinning flight: dynamics of Frisbees, boomerangs, samaras, and skipping stones*, Springer 2006.

[21] J.E. Marsden, T.S. Ratiu, *Introduction to mechanics and symmetry*, Springer 1999.

[22] I. Niven, *Irrational Numbers*, Wiley, 1956, p. 41.

[23] Y. G. Sinai, *On the foundations of the ergodic hypothesis for a dynamical system of statistical mechanics*. Sov. Math Dokl. 4 (1963) 1818–1822

[24] Y. G. Sinai, *Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards*. Uspehi Mat. Nauk 25 (1970), no. 2 (152), 141-192.

[25] Y. G. Sinai, ed. *Dynamical Systems, Ergodic Theory and Applications*. Encyclopedia of Mathematical Sciences, V.100, Springer 1999.

[26] J. Stillwell, 2003: *Elements of number theory*. Springer, 256 pp.

[27] S. Tabachnikov, *Billiards*, in Panoramas et Synthèses 1, Société Mathématique de France, 1995.

[28] T. Tokieda, *Spinning bodies: a tutorial*, in Dynamics of Extended Bodies and of the Rings, J. Souchay, ed., Lect. Notes Phys. 682, 2006.

[29] M. Wojtkowski, *The system of two spinning disks in the torus*. Physica D 71 (1994) 430-439.

Appendix A

General Billiards (1.03)

```
import copy

#constants
m=1 #positions
n=1 #velocities
eps=0.000000000001 #approximation error
res=21 #circle resolution
large=1000
eps2=.001

#rough collisions uniform distribution constants
a=1/3
b=2*sqrt(2)/3a=1/3
b=2*sqrt(2)/3

#position and velocity matrices
P0=matrix(RR,m,n)
V0=matrix(RR,m,n)
P1=matrix(RR,m,n)
V1=matrix(RR,m,n)
P2=matrix(RR,m,n)
V2=matrix(RR,m,n)

def rot(x):
    return matrix(RR,[[-1,0,0],[0,cos(x),-sin(x)],[0,sin(x),cos(x)]])
U=matrix(RR,[[-a,b,0],[b,a,0],[0,0,-1]]) #rough collision with upward normal
S=matrix(RR,[[1,0,0],[0,1,0],[0,0,-1]]) #specular collision with upward normal

#defaults
# coordinates [angle position, x, y]
p=[[0,0,0],[0,-.1,.1],[0,.1,-1],[0,.1,.1]]
# velocities
v=[[0,.1,1],[.1,1,2],[-.1,-1,-1],[-.1,2,-1]]

#filling matrices from position and velocities, allowing for an mXn array
def create_matrices(p,v):
```

```
    P0=matrix(RR,m,n)
    V0=matrix(RR,m,n)
    P1=matrix(RR,m,n)
    V1=matrix(RR,m,n)
    P2=matrix(RR,m,n)
    V2=matrix(RR,m,n)

    for i in range(0,m):
        for j in range(0,n):
            P0[i,j]=p[i][0]

    for i in range(0,m):
        for j in range(0,n):
            P1[i,j]=p[i][1]

    for i in range(0,m):
        for j in range(0,n):
            P2[i,j]=p[i][2]

    for i in range(0,m):
        for j in range(0,n):
            V0[i,j]=v[j][0]

    for i in range(0,m):
        for j in range(0,n):
            V1[i,j]=v[j][1]

    for i in range(0,m):
        for j in range(0,n):
            V2[i,j]=v[j][2]

    return (P0,P1,P2,V0,V1,V2)

#default graph
graph=Graphics()

# Catalogue of boundaries
# wall_segment= [ starting x, starting y, ending x, ending y, starting vector x,
# starting vector y, 1=line 0=circle, 0=not end 1=end, number of component,
# 0=smooth 1=rough (,associated wall)]

easy_walls=[[1,1,-1,1,-1,0,1,0,1,0],[-1,1,-1,-1,0,-1,1,0,1,0],[-1,-1,1,-1,1,0,1,0,1,0],
[1,-1,1,1,0,1,1,0,1,0]]
```

```
rough_square=[[1,1,-1,1,-1,0,1,0,1,1],[-1,1,-1,-1,0,-1,1,0,1,1],[-1,-1,1,-1,1,0,1,0,1,1]
[1,-1,1,1,0,1,1,0,1,1]]
pretty_easy_walls=[[0,1,-1,0,-1,-1,1,0,1,0],[-1,0,0,-1,0,-1,0,0,1,0],
[0,-1,1,0,1,1,1,0,1,0],[1,0,0,1,-1,1,1,0,1,0]]
medium_walls=[[0,1,-1,0,-1,-1,1,0,1,0],[-1,0,0,-1,0,-1,0,0,1,0],[0,-1,1,0,1,1,1,0,1,0],
[1,0,0,1,-1,1,1,0,1,0],[.5,.2,-.5,-.2,0,1,1,0,2,0],[-.5,-.2,.5,.2,-.2,.5,0,0,2,0]]
medium_walls2=[[0,1,-1,0,-1,-1,1,0,1,0],[-.5,-.2,.5,.2,-.2,.5,0,0,2,0],
[-1,0,0,-1,0,-1,0,0,1,0],[0,-1,1,0,1,1,1,0,1,0],[1,0,0,1,-1,1,1,0,1,0],[.5,.2,-.5,-.2,0,
astroid=[[0,1,-1,0,0,-1,0,0,1,0],[-1,0,0,-1,1,0,0,0,1,0],[0,-1,1,0,0,1,0,0,1,0],
[1,0,0,1,-1,0,0,0,1,0]]
astroid2=[[-.1,1,-1,.1,0,-1,0,0,1,0],[-1,-.1,-.1,-1,1,0,0,0,1,0],[.1,-1,1,-.1,0,1,0,0,1,
[1,.1,.1,1,-1,0,0,0,1,0]]

astroid3=[[.1,1,-.1,1,-1,0,1,0,1,0],
[-.1,1,-1,.1,0,-1,0,0,1,0],
[-1,.1,-1,-.1,0,-1,1,0,1,0],
[-1,-.1,-.1,-1,1,0,0,0,1,0],
[-.1,-1,.1,-1,1,0,1,0,1,0],
[.1,-1,1,-.1,0,1,0,0,1,0],
[1,-.1,1,.1,0,1,1,0,1,0],
[1,.1,.1,1,-1,0,0,0,1,0]
]

astroid4=[[.1,1,-.1,1,-1,0,1,0,1,1],
[-.1,1,-1,.1,0,-1,0,0,1,1],
[-1,.1,-1,-.1,0,-1,1,0,1,1],
[-1,-.1,-.1,-1,1,0,0,0,1,1],
[-.1,-1,.1,-1,1,0,1,0,1,1],
[.1,-1,1,-.1,0,1,0,0,1,1],
[1,-.1,1,.1,0,1,1,0,1,1],
[1,.1,.1,1,-1,0,0,0,1,1]
]

stadium=[[1,1,-1,1,-1,0,1,0,1,1],
[-1,1,-1,-1,-1,0,0,0,1,1],
[-1,-1,1,-1,1,0,1,0,1,1],
[1,-1,1,1,1,0,0,0,1,1]]

starsmooth=[[0,1,-.3,.3,-.3,-.7,1,0,1,0],
[-.3,.3,-1,0,-.7,-.3,1,0,1,0],
[-1,0,-.3,-.3,.7,-.3,1,0,1,0],
[-.3,-.3,0,-1,.3,-.7,1,0,1,0],
[0,-1,.3,-.3,.3,.7,1,0,1,0],
```

```
[.3,-.3,1,0,.7,.3,1,0,1,0],
[1,0,.3,.3,-.7,.3,1,0,1,0],
[.3,.3,0,1,-.3,.7,1,0,1,0] ]

starrough=[[0,1,-.3,.3,-.3,-.7,1,0,1,1],
[-.3,.3,-1,0,-.7,-.3,1,0,1,1],
[-1,0,-.3,-.3,.7,-.3,1,0,1,1],
[-.3,-.3,0,-1,.3,-.7,1,0,1,1],
[0,-1,.3,-.3,.3,.7,1,0,1,1],
[.3,-.3,1,0,.7,.3,1,0,1,1],
[1,0,.3,.3,-.7,.3,1,0,1,1],
[.3,.3,0,1,-.3,.7,1,0,1,1] ]

partrough=[[0,1,-.3,.3,-.3,-.7,1,0,1,0],
[-.3,.3,-1,0,-.7,-.3,1,0,1,0],
[-1,0,-.3,-.3,.7,-.3,1,0,1,0],
[-.3,-.3,0,-1,.3,-.7,1,0,1,0],
[0,-1,.3,-.3,.3,.7,1,0,1,0],
[.3,-.3,1,0,.7,.3,1,0,1,1],
[1,0,.3,.3,-.7,.3,1,0,1,1],
[.3,.3,0,1,-.3,.7,1,0,1,0] ]

roughstar=[[0,1,-.3,.3,-.3,-.7,1,0,1,1],
[-.3,.3,-1,0,-.7,-.3,1,0,1,1],
[-1,0,-.3,-.3,.7,-.3,1,0,1,1],
[-.3,-.3,0,-1,.3,-.7,1,0,1,1],
[0,-1,.3,-.3,.3,.7,1,0,1,1],
[.3,-.3,1,0,.7,.3,1,0,1,1],
[1,0,.3,.3,-.7,.3,1,0,1,1],
[.3,.3,0,1,-.3,.7,1,0,1,1] ]

triangle=[[1,1,-1,1,-1,0,1,0,1,0],
[-1,1,0,-1,1,-2,1,0,1,0],
[0,-1,1,1,1,2,1,0,1,0]]

rough_triangle=[[1,1,-1,1,-1,0,1,0,1,1],
[-1,1,0,-1,1,-2,1,0,1,1],
[0,-1,1,1,1,2,1,0,1,1]]

circle=[[1,0,-1,0,0,1,0,0,1,0],[-1,0,1,0,0,-1,0,0,1,0]]
circle2=[[1,0,0,1,0,1,0,0,1,0],[0,1,-1,0,-1,0,0,0,1,0],
[-1,0,0,-1,0,-1,0,0,1,0], [0,-1,1,0,1,0,0,0,1,0]]
```

```
circlesr=[[1,0,0,1,0,1,0,0,1,1],[0,1,-1,0,-1,0,0,0,1,0],
[-1,0,0,-1,0,-1,0,0,1,1], [0,-1,1,0,1,0,0,0,1,0]]

delta=.1
stadiumdelta=[[delta,1,-delta,1,-1,0,1,0,1,1],[-delta,1,-1-delta,0,-1,0,0,0,1,1], [-1-de
[delta,-1,1+delta,0,1,0,0,0,1,1], [1+delta,0,delta,1,0,1,0,0,1,1]]

# basic functions concerning wall segments
def qb(wall_segment):
    return (wall_segment[0],wall_segment[1])
def qe(wall_segment):
    return (wall_segment[2],wall_segment[3])
def diff(u1,u2):
    return (u2[0]-u1[0],u2[1]-u1[1])
def veclen(u):
    return (u[0]^2+u[1]^2)^.5
def innprod(u1,u2):
    return u1[0]*u2[0]+u1[1]*u2[1]

# this returns the (smallest) angle between two vectors
def vecang(u1,u2):
    return RR(arccos(innprod(u1,u2)/(veclen(u1)*veclen(u2))))

# print(vecang((0,1),(0,-1)))

def ub(wall_segment):
    w=(wall_segment[4],wall_segment[5])
    l=veclen(w)
    return (wall_segment[4]/l,wall_segment[5]/l)

# Testing simple functions above:
# show(veclen(diff(qb(easy_walls[1]),qe(easy_walls[1]))))

# normal vector from qb and qe
def normal(u1,u2):
    w=diff(u1,u2)
    l=veclen(w)
    return (-w[1]/l,w[0]/l)

#show(normal(qb(easy_walls[1]),qe(easy_walls[1])))

def curv(u1,u2,u3):
    # u1=qb, u2=qe, u3=ub
```

```
    norm=normal(u1,u2)
    return 2*(u3[0]*norm[0]+u3[1]*norm[1])/veclen(diff(u1,u2))

#show(curv(qb(easy_walls[1]),qe(easy_walls[1]),ub(easy_walls[1])))

def ue(u1,u2,u3):
    # u1=qb, u2=qe, u3=ub
    norm=normal(u1,u2)
    #show(norm)
    #show(u3)
    l=(2*(u3[0]*norm[0]+u3[1]*norm[1]))
    w=(norm[0]*l,norm[1]*l)
    return diff(w,u3)

#show( ue( qb(easy_walls[1]),qe(easy_walls[1]),ub(easy_walls[1])) )

def theta(u1,u2,u3,u4):
    # qb,qe,ub,ue
    if innprod(diff(u1,u2),u3)>=0:
        return arccos(innprod(u3,u4))
    if innprod(diff(u1,u2),u3)<0:
        return 2*pi-arccos(innprod(u3,u4))

# show(theta(qb(easy_walls[1]), qe(easy_walls[1]), ub(easy_walls[1]), ue(qb(easy_walls[1

def center(u1,u2,u3):
    #qb,qe,ub
    cu=curv(u1,u2,u3)+eps
    return diff((-u3[1]/cu,u3[0]/cu),u1)

#show(center(qb(easy_walls[1]), qe(easy_walls[1]), ub(easy_walls[1])))

def pdistance(u1,u2):
    return ((u1[0]-u2[0])^2+(u1[1]-u2[1])^2)^(1/2)


def getangle(A):
    # return the angle of the vector A relative to the pos x-axis
    if A[0]>=0:
        if A[1]>=0:
            return (arcsin((A[1]-eps)/veclen(A)))
        if A[1]<0:
            return (arcsin((A[1]+eps)/veclen(A)))
```

```
    else:
        #print('down here')
        return RR(pi-arcsin((A[1]-eps)/veclen(A)))

def circ_or(v,w):
    if v[0]*w[1]-v[1]*w[0]>0:
        #print('counter clockwise')
        return 0
    else:
        #print('clockwise')
        return 1

def draw_table(wall_segments):
    #print(len(wall_segments))
    newgraph=Graphics()
    for i in range(0,len(wall_segments)):
        #print('i',i)
        if wall_segments[i][6]==1:
            # draw line
            newgraph+=plot(line([(wall_segments[i][0],wall_segments[i][1]),
            (wall_segments[i][2],wall_segments[i][3])]))
        else:
            u1=qb(wall_segments[i])
            u2=qe(wall_segments[i])
            u3=ub(wall_segments[i])
            u4=ue(u1,u2,u3)
            #print(u1,u2,u3,u4)
            C=center(u1,u2,u3)
            #print('center',C)
            arc_angle=RR(theta(u1,u2,u3,u4))
            #print('total angle', arc_angle)
            radius=1/abs(curv(u1,u2,u3))
            #print('radius', radius)
            x0=u1[0]
            y0=u1[1]
            startvec=diff(C,u1)
            #print('start vector',startvec)
            alpha=getangle(startvec)
            #print('alpha',alpha)
            for j in range(0,res+1):
                omega=(-1)^(circ_or(u3,diff(u1,u2)))*j/res*arc_angle
                #print(omega)
                #print(alpha)
```

```
                #print(radius*cos(alpha+omega),radius*sin(alpha+omega))
                #print(j,alpha+omega)
                x1=radius*cos(alpha+omega)+C[0]
                y1=radius*sin(alpha+omega)+C[1]
                #print(x1,y1)
                L=line([(x0,y0),(x1,y1)])
                newgraph=newgraph+plot(L)
                x0=x1
                y0=y1
                #show(newgraph)
    #show(newgraph)
    return newgraph


def pointonarc(x,y,wall):
    u1=qb(wall)
    u2=qe(wall)
    u3=ub(wall)
    u4=ue(u1,u2,u3)
    C=center(u1,u2,u3)
    thet=theta(u1,u2,u3,u4)
    v1=diff(C,(x,y))
    v2=diff(C,u1)
    ori=circ_or(diff(C,u1),diff(C,u2))
    #print('ori',ori)
    ang=vecang(v1,v2)
    #print('theta', thet, 'ang',ang)
    if circ_or(v1,v2)==0:
        angdis=2*pi-ang
    else:
        angdis=ang
    if ori==1 and (thet>pi+eps or thet<pi-eps):
        if thet>2*pi-angdis:
            #print('1')
            return True
        if thet<2*pi-angdis:
            return False
    if ori==0 and (thet>pi+eps or thet<pi-eps):
        if thet>angdis:
            #print('2')
            return True
        if thet<angdis:
            return False
```

```
    if thet==pi:
        if innprod(v1,u3)>0:
            return True
        else:
            return False


def wall_collide(pos,vel,wall):
    #print('in collide')

    # first check if the trajectory intersects the segment for linear boundaries
    # Note: works for only some circles--need better solution
    # It will excluse some legitimate circles if the trajectory passes in and out the ar
    if wall[6]==1:
        v1=diff(pos,(wall[0],wall[1]))
        v2=diff(pos,(wall[2],wall[3]))
        #print(v1,v2,(vel[1],vel[2]))
        #print(vecang((vel[1],vel[2]),v1),vecang((vel[1],vel[2]),v2),vecang(v1,v2))
        #print(vecang((vel[1],vel[2]),v1)+vecang((vel[1],vel[2]),v2))
        angle_difference=vecang((vel[0],vel[1]),v1)+vecang((vel[0],vel[1]),v2)-vecang(v1
        #note cannot use vel directly as it is 3D
        #print(angle_difference)
        if angle_difference>0.001:
            #print('does not collide')
            return (0,0,large+1)

    # m2 is the slope of vel, handling infinite case by making it large
    if vel[0]==0:
            m2=1000000000
    else:
            m2=vel[1]/vel[0]

    if wall[6]==0:
        #print('checking arc')
        u1=qb(wall)
        u2=qe(wall)
        u3=ub(wall)
        C=center(u1,u2,u3)
        radi=pdistance(C,u1)
        #print('radius',radi)
        A=1+m2^2
        B=-2*C[0]+2*m2*(pos[1]-m2*pos[0]-C[1])
        c=C[0]^2+m2^2*pos[0]^2+2*m2*pos[0]*(C[1]-pos[1])+(pos[1]-C[1])^2-radi^2
```

```
    x1=(-B+(B^2-4*A*c)^(1/2))/(2*A)
    x2=(-B-(B^2-4*A*c)^(1/2))/(2*A)
    # This handles the cases where the trajectory misses the circle entirely
    if x1.imag()<>0:
        return (0,0,large+1)

    #if x2.real==False:
    #    return (0,0,large+1)
    else:
        y1=pos[1]+m2*(x1-pos[0])
        y2=pos[1]+m2*(x2-pos[0])
        #print(x1,y1,pointonarc(x1,y1,wall))
        d2=pdistance(pos,(x2,y2))
        d1=pdistance(pos,(x1,y1))
        #print(d1,d2)
        if pointonarc(x1,y1,wall):
            #print('made it here')
            pointonarc(x2,y2,wall)==False,  (vel[0]<0 and x2-pos[0]>0) or (vel[0]>0
            if d1>eps and ((vel[0]>0 and x1-pos[0]>0) or (vel[0]<0 and x1-pos[0]<0))
                #print('returning x1')
                return (x1,y1,d1)
        #print(x2,y2,pointonarc(x2,y2,wall))
        if pointonarc(x2,y2,wall):
            #print('made it to x2 part')
            #print(d2>eps,d2,eps)
            if d2>eps and ((vel[0]>0 and x2-pos[0]>0) or (vel[0]<0 and x2-pos[0]<0))
                #print('returning x2')
                return (x2,y2,d2)
        #print(x1,y1,d1)
        #print(x2,y2,d2)
    return(0,0,large+1)


if wall[6]==1:
    # find the intersection point
    if wall[2]==wall[0]:
        m1=(wall[3]-wall[1])/(eps)
    else:
        m1=(wall[3]-wall[1])/(wall[2]-wall[0])

    if m2==m1:
        m2=m2-eps
    x=(m2*pos[0]-pos[1]-m1*wall[0]+wall[1])/(m2-m1)
```

```
        y=pos[1]+m2*(x-pos[0])
        if ((x-pos[0])^2+(y-pos[1])^2)<eps:
            #print('not supposed to be here')
            return (0,0,large+1)
        else:
            #print('found a collision')
            return (x,y,((x-pos[0])^2+(y-pos[1])^2))


def next_wall(walls,rotational_position,x,y,rotational_velocity,vx,vy):
    #print('in next wall')
    best_dist=large
    bestx=x
    besty=y
    best_wall=0
    for i in range(0,len(walls)):
        #print('wall', i)
        #note: dist=distance squared
        (xnew,ynew,dist)=wall_collide((x,y),(vx,vy),walls[i])
        #print('dist',dist)
        if eps<dist<best_dist:
            bestx=xnew
            besty=ynew
            best_dist=dist
            best_wall=i
    return ((best_dist/(vx^2+vy^2)^(1/2))*rotational_velocity+rotational_position,bestx,

def new_positions(walls,M0,M1,M2,N0,N1,N2):
    #print('new positions')
    Mw=matrix(QQ,m,n)
    for i in range(0,m):
        #print(i)
        for j in range(0,n):
            #$print(j)
            #show(M1)
            (M0[i,j],M1[i,j],M2[i,j],Mw[i,j])=next_wall(walls,M0[i,j],M1[i,j],M2[i,j],N0

        #show(Ptemp)

    return (M0,M1,M2,Mw)

def get_tangent(wall,x,y):
    if wall[6]==1:
```

```
        return getangle(diff(qb(wall),qe(wall)))
    if wall[6]==0:
        #print('circ tan')
        u1=qb(wall)
        u2=qe(wall)
        u3=ub(wall)
        C=center(u1,u2,u3)
        perp=getangle(diff(C,(x,y)))
        return perp+pi/2
    else:
        return False

def reflect(wall,x,y,vr,vx,vy):
    V=matrix(RR,3,1)
    V[0,0]=vr
    V[1,0]=vx
    V[2,0]=vy
    #show(V)
    # omega is the angle of the line or tangent to the circle, oriented by the direction
    omega=get_tangent(wall,x,y)
    U1=rot(omega)
    U2=rot(-omega)
    if wall[9]==0:
        #print('here thu')
        #show(U1,U2)
        T=matrix(RR,n,n)
        T=U1*S*U2
        #show(V)
        V=T*V
    else:
        V=rot(omega)*U*rot(-omega)*V
    #show(V)
    return (V[0,0],V[1,0],V[2,0])

def new_velocities(walls,P1,P2,V0,V1,V2,Pw):
    for i in range(0,m):
        for j in range(0,n):
            (V0[i,j],V1[i,j],V2[i,j])=reflect(walls[Pw[i,j]],P1[i,j],P2[i,j],V0[i,j],V1[
    return (V0,V1,V2)

def Draw_paths(X1,Y1,X2,Y2):
    paths=Graphics()
    #show(X1,X2,Y1,Y2)
```

```
    for i in range(0,m):
        for j in range(0,n):
            #print((O[i][j][1],O[i][j][2]),(N[i][j][1],N[i][j][2]))
            paths+=plot(line([(X1[i,j],Y1[i,j]),(X2[i,j],Y2[i,j])]))
            #show(paths)
    return paths

# This tests Draw_paths
# P1=[[[1,1,1],[1,1,1],[1,1,1]],[[1,1,1],[1,1,1],[1,1,1]],[[1,1,1],[1,1,1],[1,1,1]]]
# P2=[[[1,2,2],[1,0,0],[1,2,4]],[[1,12,12],[1,12,12],[1,11,1]],[[1,11,1],[1,11,1],[1,11,
# testgraph=Draw_paths(P1,P2)
# show(testgraph)

# wall_segments is a k by 10 matrix, Renato's 9 plus the tenth: 0=smooth, 1=rough
# N_max is the maximum iterations before stopping
# draw
def aalpha(x):
    return RR(arccos(32/9*(cos(x))^4-16/3*(cos(x))^2+1))

def Billiard(wall_segments, N_max, draw, p, v, theta):
    print(RR(pi/2+aalpha(theta)/2))
    if draw==1:
        graph=draw_table(wall_segments)
    (P0,P1,P2,V0,V1,V2)=create_matrices(p,v)
    graphtemp=Graphics()
    graphtemp1=Graphics()
    graphtemp2=Graphics()
    v1=1
    x0=5
    r0=0
    pl=[]
    su=0
    su2=0
    count1=0
    count2=0
    maxx=0
    for i in range(0,N_max):
        v0=v1
        v1=abs(V2[0,0])
        #print('loop',i)
        OP1=P1[:]
        OP2=P2[:]
        or0=r0
```

```
r0=P0[0,0]
#print(r0)
ox0=x0
x0=P1[0,0]
y0=P2[0,0]
#print(x0)
r0p=RR((2^.5*sin(theta/2)*y0+r0)/(1+2*(sin(theta/2))^2))
y0p=RR((2^.5*sin(theta/2)*(2^.5*sin(theta/2)*y0+r0))/(1+2*(sin(theta/2))^2))
#r0ps=RR(r0p/(x0*tan(theta/2)))
#if r0p>0:
#    par=0
#else:
#    par=1
#r0ps=RR((r0p^2+y0p^2)^.5*(-1)^par/(x0*tan(theta/2)/v0))
#print(r0,r0p)
#print(x0,x0p)
#show(V1,V2)
(P0,P1,P2,Pwall)=new_positions(wall_segments,P0,P1,P2,V0,V1,V2)
#print('made')
rv0=V0[0,0]
xv0=V1[0,0]
yv0=V2[0,0]
rvp0=RR((2^.5*sin(theta/2)*yv0+rv0)/(1+2*(sin(theta/2))^2))
(V0,V1,V2)=new_velocities(wall_segments,P1,P2,V0,V1,V2,Pwall)
rv1=V0[0,0]
xv1=V1[0,0]
yv1=V2[0,0]
rvp1=RR((2^.5*sin(theta/2)*yv1+rv1)/(1+2*(sin(theta/2))^2))


pathgraph=Draw_paths(OP1,OP2,P1,P2)
#show(pathgraph)
r1=P0[0,0]
x1=P1[0,0]
if x1>x0:
    col='red'
else:
    col='blue'
y1=P2[0,0]
r1p=RR((2^.5*sin(theta/2)*y1+r1)/(1+2*(sin(theta/2))^2))
#r1ps=RR(r1p/(x1*tan(theta/2)))
y1p=-RR((2^.5*sin(theta/2)*(2^.5*sin(theta/2)*y1+r1))/(1+2*(sin(theta/2))^2))
#if r1p>0:
```

```
#    par=0
#else:
#    par=1
#r1ps=RR((r1p^2+y1p^2)^.5*(-1)^par/(x1*tan(theta/2)/v1))
lineover=line([(x1,r1),(x0,r0)], thickness=.3, dpi=200,color=col)
PP0=point((x0,r0),size=40)
PP1=point((x1,r1),size=40)

lineover1=line([(x1,r1p),(x0,r0p)], thickness=.3, dpi=200)
PP0p=point((x0,r0p),size=40)
PP1p=point((x1,r1p),size=40)
graph=graph+pathgraph
if x1>maxx:
    if (-1)^i>0:
        maxang=getangle([xv1,rvp1])
    else:
        maxang=getangle([xv1,-rvp1])
    maxx=x1
if i>0:
    #print(RR(xv1^2+rvp1^2))
    if (-1)^i>0:
        pl+=[[getangle([xv1,rvp1]),x1]]
    else:
        pl+=[[RR(getangle([xv1,-rvp1])),x1]]
    #pl+=[[xv1,rvp1],[xv0,rvp0]]
    #pl+=[[i,(getangle([xv1,rvp1])+getangle([xv0,rvp0]))/2]]
if x1>5 and x0>5:
    if r1>r0:
        count1+=1
        tem=(r1-or0)/(x1-ox0)
        su+=tem
    if r1<r0:
        count2+=1
        tem2=(r1-or0)/(x1-ox0)
        su2+=tem2

    graphtemp=graphtemp+plot(lineover)+plot(PP0)+plot(PP1)
if i>0:
    graphtemp1=graphtemp1+plot(lineover1)+plot(PP0p)+plot(PP1p)
#lineover2=line([(x1,r1ps),(x0,r0ps)], thickness=.3, dpi=200)
#PP0ps=point((x0,r0ps),size=40)
#PP1ps=point((x1,r1ps),size=40)
```

```
        #if i>0:
        #    graphtemp2=graphtemp2+plot(lineover2)+plot(PP0ps)+plot(PP1ps)
        if OP1==P1 and OP2==P2:
            print('error')
            print(i)
            #return
        #if i>14:
        #    print('P1',P1)
        #    print(P2)
        #    print(V1)
        #    print(V2)
        #print(P0)
    #P=NewP
    #V=NewV
    #print(2*su/N_max)
    #print(RR(2*su/N_max/theta))
    #pl+=[[RR(theta),2*su/count1],[RR(theta),2*su2/count2]]
    print(maxang,maxx,aalpha(theta))
    pl2=[[theta,maxang]]
    return (P0,P1,P2,V0,V1,V2,graph,graphtemp,graphtemp1,pl)


graph=Graphics()
graphlist=[]
def triangle(theta,s):
    y=RR(s*tan(theta/2))
    return [[s,y,0,0,-s,-y,1,0,1,1],[0,0,s,-y,s,-y,1,0,1,1],[s,-y,s,y,0,1,1,0,1,1]]
for i in range(0,10):
    print(i)
    graph=Graphics()
    graphpr=Graphics()
    graphax=Graphics()
    graphax2=Graphics()
    table=triangle(.01+pi/25*i,1+5*(i+1))
    p=[[0,5,0]]
    v=[[-.1,.1,RR((1-.1^2*2)^.5)]]
    (P0,P1,P2,V0,V1,V2,graph,graphpr,graphax2,pl)=Billiard(table, 500, 1, p, v,.01+pi/25
    graphlist+=pl
    #graph.show(aspect_ratio=True,axes=False)
    #graphpr.show(axes=1, figsize=10)
    #graphax.show(axes=False)
    #graphax2.show(axes=False)
graph2=list_plot(graphlist,size=5,color='black')
```

```
graph2.show(axes=1, figsize=5)
```